2010

Investigating the impact of segmented and whole-text repetition on listening comprehension, comprehension processes, and comprehension problems

Anne O'bryan
Iowa State University

Investigating the impact of segmented and whole-text repetition on listening comprehension, comprehension processes, and comprehension problems

by

Anne O'Bryan

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Applied Linguistics and Technology

Program of Study Committee:
Volker Hegelheimer, Major Professor
Carol Chapelle
John Levis
Lee Honeycutt
Alison Morris
Mack Shelley

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Thanks to everyone who helped me complete this doctoral program and, ultimately, this dissertation. First, thanks to Dr. Volker Hegelheimer for his guidance and support throughout my coursework, research, and writing of this dissertation. Volker, I appreciate your encouragement, enthusiasm, and willingness to listen as I talk through ideas. I am grateful for the support of my committee members who provided comments and questions along the way, and for encouraging me to pursue my areas of interest. Thanks in general to the Applied Linguistics and Technology faculty for their enthusiasm and wisdom throughout the seven years I've been a student at Iowa State.

I want to especially thank my friends and family for encouraging me and making it possible for me to be a full-time scholar. Patrick and Quinlan, I couldn't have done this without your love and support. Thanks to my parents, sister, in-laws, and friends for babysitting when I had work to do, and thanks to my friends and colleagues who served as extra raters and sometimes guinea pigs during various stages of my research. I am indebted to you all.

ABSTRACT

This dissertation presents a product- and process-oriented approach to investigating how a common teaching methods used in second language listening classrooms—repetition—might be structured in a way that enhances listening comprehension and facilitates input processing. The major purpose of the dissertation was to investigate the impact of both whole-text and semantically segmented repetition on English as a Second Language (ESL) learners' listening comprehension and comprehension problems.

An embedded, mixed-methods approach was employed in the study. Quantitative data consisted of written recalls from 100 intermediate-level, ESL participants who each listened to two, authentic lecture excerpts in one of the two aforementioned conditions. Qualitative data consisted of verbal protocols and post-task interviews with 12 participants.

The findings indicate that there was no statistically significant difference between overall recall scores (i.e., summed recall scores for the two texts) for students in the two conditions. When recall scores for each lecture were compared individually, the results mirrored those found when looking at the scores overall in terms of statistical significance. Despite the lack of statistical significance from the recalls, findings from the verbal protocol data largely supported the hypotheses that students in the whole-text condition would encounter more problems attending to the task and forming a mental representation of the input compared to students in the segmented condition who were

provided with more guidance in attending to the task and structure building through the task's design.

The results from the study were mixed in terms of supporting and refuting the claims in previous literature. However, both the findings and methods from the study hold a number of implications and recommendations for language teachers, materials developers, those interested in the use of technology for language learning and assessment, and future researchers.

CHAPTER 1: INTRODUCTION

Statement of the problem

Although each of the four skills (listening, reading, writing and speaking) is key for second language (L2) acquisition, listening is viewed as the "primary means of L2 acquisition " (Rost, 2002, p. 103). Despite this, Vandergrift (2007) notes that "L2 listening remains the least researched of all four language skills" (p. 191). Most L2 listening research has investigated either the *product* of listening by examining what effect a listening condition (e.g., exposure to different types of tasks, different pre-listening activities, etc.) has on overall comprehension, or taken a closer look at the cognitive processes employed by listeners while listening (i.e., the *process* of listening) (Vandergrift, 2007). While the relatively small body of L2 listening research has informed teaching techniques, much of the focus remains on the product of listening. Teachers, therefore, might know what impact a certain task variable would have on students' comprehension scores, but at the same time have no idea *how* students arrived at a certain answer or what problems they encountered as they engaged in the listening task.

Combining product- and process-based research can help classroom instructors gain a deeper understanding of how common tasks advocated in listening textbooks affect L2 listeners. Specifically, repetition is often encouraged in L2 listening textbooks (e.g. Dunkel, Pialorsi, and Kozyrev, 1996; Lebauer, 2000; Salehzadeh, 2006). A number of researchers have conducted product-based research on the impact of this variable on

overall comprehension and found that repetition does, indeed, lead to an increase in listening comprehension scores (see Cabrera and Martinez, 2001; Cervantes and Gainer, 1992; Jensen and Vinther, 2003; and O'Bryan and Hegelheimer, 2009, for examples). However, little is known about the processes listeners employ—the strategies they use, or the problems they encounter—while engaging in this type of task.

Empirical work in repetition and listening comprehension does little to direct teachers in the ways of providing repeated input that will reap the most benefits, e.g., positively influence listening comprehension processes and increase overall listening comprehension. In looking at ways to address this problem, researchers in the areas of both psychology and applied linguistics (see Samuels and LaBerge,1983, and van Dijk, 1981, respectively) suggest that dividing longer, more complex texts into shorter, thematically-based units may ease the cognitive burden caused by both time on task and task complexity and facilitate comprehension. Therefore, combining the segmenting of long texts into shorter, thematically- or semantically-based units with the provision of repetition can theoretically help L2 listeners overcome some of the common problems noted in L2 listening comprehension research (see Bacon, 1992; Goh, 2000) and create a structured representation of the text in memory, thereby enhancing recall.

In sum, regardless of the existing research investigating the impact of repetition on L2 listening comprehension, there is a glaring omission of empirical work investigating the *structuring* of repeated input, specifically using semantic text chunking, on L2 listening

comprehension and comprehension processes despite theoretical support in the areas of both psychology and applied linguistics claiming that this type of task modification could ease the learner's cognitive burden and facilitate comprehension. Dunkel (1991) calls for teachers and researchers "to increase vastly the number of empirical studies that investigate the ways in which factors inside and outside the L2 head affect comprehension of L2 discourse for beginning-, intermediate- and advanced-level L2 listeners" (p. 445). The present study addressed Dunkel's concern by investigating the ways in which "factors inside and outside the L2 head" (p. 445), such as comprehension problems, processes, and task type, influence listening comprehension for intermediate-proficiency learners of English as a Second Language (ESL).

Purpose of the study

The purpose of this mixed methods study (Creswell, 2003) was to explore the impact of two types of listening tasks, namely the offering of repeated input in either a whole-text or segmented format, on written recalls, which were used to measure comprehension. In addition, this study explored the impact of these task types on the listening comprehension problems students experienced in an attempt to determine whether one task type was more effective at helping learners overcome common problems (e.g., missing information, being unable to form a mental representation of the text from words heard, etc.) identified in previous literature (see Bacon, 1992; Goh, 2000).

Significance of results

The knowledge gained from this study has implications for language teachers, materials developers, and the use of technology in language learning and assessment. First, having a more complete understanding of the way in which learners process information, as well as the difficulties they encounter during the three phrases of comprehension (Anderson, 1985), can help instructors learn to tailor instruction to comprehension problems that are commonly encountered, as well as adapt and structure materials in ways that facilitate input processing in order to enhance comprehension. Materials developers can also benefit from the study, as results of the study can contribute to the necessary preparatory work that is essential to designing useful, relevant materials. Just as there are implications for teachers and materials developers, there are also implications stemming from both the methods and results of the present study for the use of technology in language learning and assessment by both teachers and researchers.

Organization of the dissertation

This dissertation is divided into five chapters. In chapter one I presented the purpose of the study and the research objectives. In chapter two, I review relevant literature in the areas of listening processes and comprehension, input processing, and repetition, and finally discuss how attention and structure building can be enhanced through segmented repetition. Chapter three contains a description of the methods, including materials,

measures, procedure, and analysis of the research. I present the results of the research in chapter four along with a discussion of these findings. Finally, Chapter five discusses limitations and implications of the research and makes suggestions and recommendations in light of the results.

CHAPTER 2: LITERATURE REVIEW

This chapter is a literature review that elaborates on the areas of listening processes and comprehension, input processing, and repetition, and finally discusses how attention and structure building can be enhanced through segmented repetition. The theoretical and empirical background informing this study combines perspectives from the areas of cognitive psychology, discourse studies, and applied linguists. The interdisciplinary nature of this study highlights the complexity one encounters when conducting process-based listening research, yet at the same time, some essential insights from these areas have strong links to current and potential pedagogical practice. First, repetition as a task variable has received considerable attention in the teaching of L2 listening due to its ability to provide learners with more time to process input for both meaning and form. In addition, repetition and the division of tasks into smaller units have been shown to raise understandability. However, specifics on how these smaller units should be identified has not been provided to L2 listening practitioners.

This chapter introduces the notion that repetition, along with the division of texts into thematically-based units, could help L2 listeners create a structured representation of the text in memory and enhance recall. Making new topics salient and their content expectable increases learners' chances of understanding, while semantic chunking can facilitate parsing, a crucial phase in the listening comprehension process. By enhancing attention and structure building through structured repetition, it is conceivable that this type of task condition could help students overcome some of the listening comprehension problems identified in published research.

After reviewing the literature and highlighting key gaps, which provide strong support for the topic of this dissertation work, a set of research questions stemming from this review is presented.

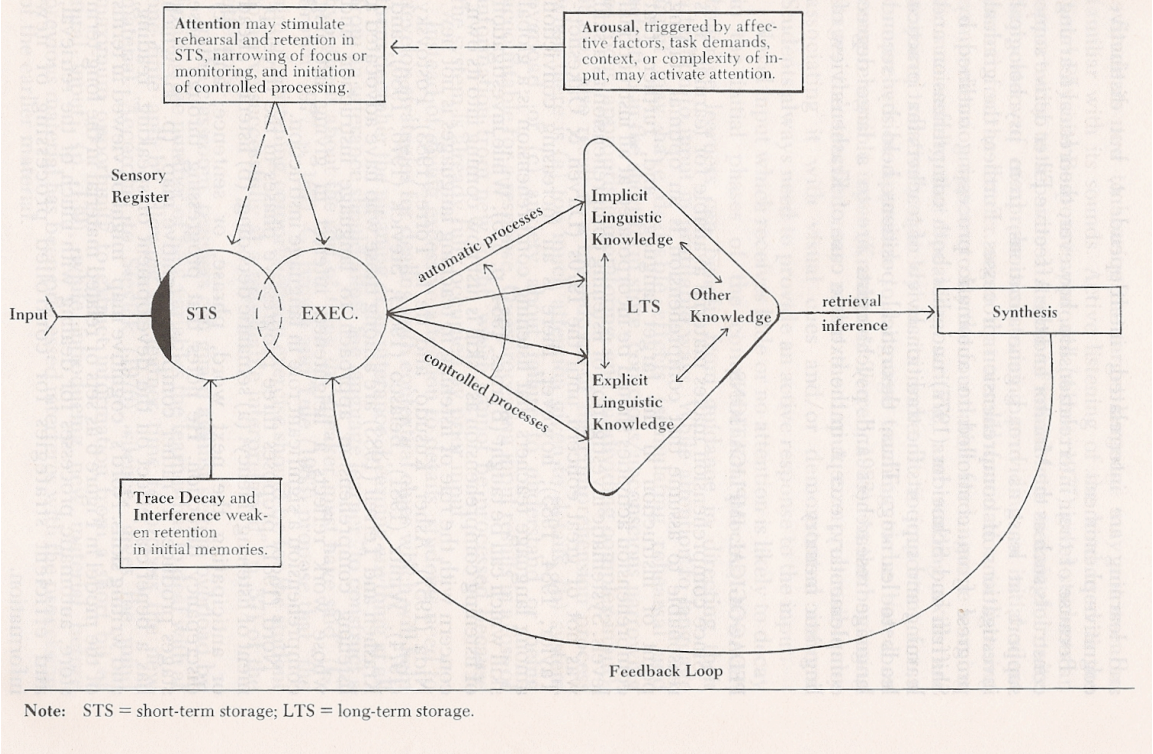
Models of listening comprehension

The process of listening in a second or foreign language is quite complex and is often a source of frustration for learners (Graham, 2006). As Rost (2005) notes,"[l]istening encompasses receptive, constructive, and interpretive aspects of cognition, which are utilized in both first language (L1) and second language (L2) listening" (p. 503). Listening involves bottom-up processing, in which listeners use their linguistic knowledge of sounds, word forms and grammatical relationships to comprehend input, as well as top-down processing, where prior experience, real-world knowledge or familiarity with the listening context help the listeners to interpret an utterance (Peterson, 2001). These processes can take place simultaneously, though this is more frequently found with higher-proficiency listeners (O'Malley, Chamot, and Küpper, 1989; Peterson, 2001).

In studying listening comprehension, many researchers focus on input processing based on Anderson's (1985) three-phase comprehension model: perceptual processing, parsing and utilization. Anderson notes that these phases are ordered, by necessity, in time but also partly overlap. In the perceptual stage, the learner recognizes sounds and segments those sounds into words. These words are accessed by various clues, including the identification of phonemes and recognition of word boundaries and syllable stress (Rost, 2002). In the parsing stage, listeners assign recognized words into grammatical categories and assign structural and semantic relations. These words are then transformed into a

mental representation of the combined meaning of words; this information is moved to long-term memory and stored as propositions. Once a sentence or utterance has been parsed and mapped into a meaning representation, learners in the utilization stage begin making connections between this newly-parsed information and the knowledge they have about the world.

Anderson's (1985) three-phase model of language comprehension is not specific to L2, nor listening, comprehension; while the underlying processes remain the same for both L1 and L2 listeners, Nagle and Sanders' (1986) model of listening comprehension processing in the adult language learner (see Figure 1) expands on the three-phase view provided by Anderson (1985) by helping explain how and why overlap occurs among the three phases; it also provides additional insight into problems and processes encountered.



Note: STS = short-term storage; LTS = long-term storage.

Figure 1. Nagle and Sanders' (1986) model of listening comprehension. From "Comprehension theory and second language pedagogy" by S. Nagle and S. Sanders, 1986, *TESOL Quarterly, 20*(1), p. 19. Reprinted with permission.

In this model, input is processed first in the sensory register, a part of short-term memory where sounds are stored for a short time. Problems such as trace decay—a fading of the sensory input that happens about 1-2 seconds after receiving it (Field, 2004)—and interference from newly arriving input may occur in short-term storage (STS) and prevent L2 listeners from recognizing sounds and segmenting them into words, as in Anderson's perceptual processing phase. To remedy these problems, learners may need to allocate attention to specific features of input; attention may also stimulate rehearsal, or conscious or unconscious repetition of the input, which "may strengthen an item in short-term memory" (Nagle and Sanders, 1986, p. 15). Nagle and Sanders relate the role of attention in input processing to Krashen's (1977, 1982) view of the Monitor's role in language production where the "Monitor focuses on form(s); that is, it analyzes (or subdivides) linguistic units into smaller components. If one views the Monitor as an input processor as well, monitoring may be described as the directing of attention to specific input (or output) items" (Nagle and Sanders, 1986, p. 17). A major factor in activating the Monitor is arousal, "an increase of activity in the nervous system" (p. 17) which not only activates attention but also encourages appropriate controlled processing and monitoring. In Nagle and Sanders' model, the *executive* oversees processing operations and controls the flow of information.

As seen in Figure 1, the executive controls which processes are automatic, and which are controlled. Automatic processing is a "sequence of nodes that nearly always becomes active in response to a particular input configuration" (Shiffrin and Schneider, 1977, p. 155); these nodes are activated automatically, with no need for special attention or control on the part of the learner. Controlled processes use "a temporary sequence of nodes activated under control of, and through attention by, the subject" (Shiffrin and Schneider, 1977, p. 156); Nagle and Sanders (1986) note that this type of processing can be encouraged by certain types of task demands, and that it "is not necessarily conscious in all cases" (p. 16). In an L2 environment, McLaughlin, Rossman, and McLeod (1983) found that controlled processing occurs when learners interact with new language and language activities that require a great amount of attention, while most automatic processing occurs incidentally. Nagle and Sanders (1986) state that "[a]utomatic processing is critical to comprehension because too much controlled processing may lead to overload and breakdown" (p. 16). However, Vandergrift (2007) notes that beginning-level listeners automatically process "little of what they hear" (p. 193). If automatic processes are not available or activated during a comprehension task, learners must use their mental energy to *attend to* specific features of input and/or to the controlled processes used for retrieval.

Once the executive decides which stores of long-term storage (LTS), whether linguistic or other knowledge, are activated in the parsing stage (Anderson, 1985), the "retrieved knowledge [from LTS] and the individual's judgments (inference) about unfamiliar data" (p. 18) are synthesized to make meaning of the input. This is similar to Anderson's third

stage, utilization. Nagle and Sanders' model shows the results of this synthesis, even if incorrect, being fed back to the executive where reprocessing takes place, if necessary.

L2 listening research has drawn on the work of both Anderson (1985) and Nagle and Sanders (1986) to study the differences in the comprehension processes and listening strategies used by effective and ineffective listeners, identify problems that take place during the three phases of comprehension, and identify the listening strategies in which students engage while progressing through the three phases.

Researching Input Processing in L2 Listening Comprehension

Much of the literature detailing ways in which L2 listeners process input has been carried out by listening strategies researchers who have found that variables such as the effectiveness of the listener and listening proficiency level impact the strategies L2 listeners use. O'Malley et al. (1989) and Bacon (1992) used introspective and retrospective think-aloud data to show that effective and ineffective listeners use different strategies during the three phases of comprehension (Anderson, 1985). While, overall, successful listeners reported a greater number and range of strategies and were more likely to successfully combine top-down and bottom-up processing in order to comprehend (Bacon, 1992), O'Malley et al. (1989) found that during the perceptual processing stage, effective listeners consciously focused their attention on the task, whereas ineffective listeners "reported that when they encountered an unknown word or phrase…they usually just stopped listening" (p. 428). In the parsing stage, effective listeners listened for larger chunks of information and focused on individual words only when there was a breakdown in comprehension, whereas ineffective listeners tended to

use more bottom-up processes and concentrate on individual words. In the utilization stage, listeners tended to elaborate on the input by drawing on world and personal knowledge; effective listeners were found to relate new information to prior knowledge, while ineffective listeners had fewer elaborations overall, and "did not make connections between the new information and their own lives" (p. 432), a finding also reported by Bacon (1992).

As the effectiveness of the listener certainly impacts the processing strategies and difficulties faced by a listener, so too does listening proficiency level. Research into the impact of listening proficiency level on listening strategies highlights the importance of relating new information to prior knowledge (see Bacon, 1992) and emphasizes the use of predictive strategies by engaging semantic memory, a strategy employed by some of the more-skilled, non-native English listeners in O'Malley et al. (1989). Conrad (1983) administered postlistening and conventional cloze tests to ESL learners at the intermediate and advanced levels, along with native speakers, and found that "with increased proficiency in the language, listeners were found to base their responses on contextual semantic [i.e. meaning] cues from the text and to rely less on the syntactic [i.e. grammatical structure] cues" (p. 67). This is in contrast to VanPatten (2004) who, in discussing his input processing theory, notes that low-level learners "process content words in the input before anything else" (p. 8). Because beginning L2 learners have limited resources for processing input, VanPatten states "certain elements of [grammatical] form will not get processed for acquisitional purposes" (p. 7).

Although there is some disagreement about which cues predominate at different levels of proficiency, learners clearly make use of both semantic and syntactic cues during the comprehension process; it is when the learner's combination of syntactic and semantic processing are "in conflict…[that] comprehension is hurt" (Anderson, 1985, p. 347). Given the fact that "native speakers of English use primarily *semantic cues* (i.e., information provided by the context) to process aural texts" (Berne, 2004, p. 521, emphasis in original), Berne suggests that "[nonnative English listeners] should be encouraged to develop listening strategies that focus more on…semantic cues and less on syntactic cues (p. 522). Engaging semantic memory—"built up of years of experience with the world" (Kutas and Federmeier, 2000, p. 466)—means that listeners must relate new information in a listening text to information stored in their long-term memory. Doing so enables a listener to predict plausible alternatives and, therefore, "constrain the search through semantic memory and facilitate the processing of the item(s) most likely to appear" (p. 467).

While the engaging of semantic memory can facilitate input processing, learners must first possess some background knowledge of the subject from which to draw when processing new information. In discussing the importance of topic familiarity in facilitating recall, Rost (1994) notes that "[e]xperts in a topic area generally remember more of what they heard, not because of superior language or recall ability, but because of…familiarity with facts and ideas that non-experts may consider 'new information', and prior practice with drawing inferences in a particular topic area" (p. 97). After finding that topic familiarity had a significant effect on the recall of two listening texts by

90 first-, second- and third-quarter university Spanish students, Schmidt-Rinehart (1994) concluded that "helping students make connection to their previous knowledge in order to build a mental framework with which to link the new information might facilitate comprehension" (p. 185). Vandergrift and Tafaghodtari (2010) provided empirical evidence that students can be taught to make these connections. They led 59 students of French as a second language through the process of engaging background knowledge through prediction and discussion exercises both before and during a listening task and found that this process significantly influenced overall listening comprehension. Possessing background knowledge and then engaging it when listening can, then, facilitate inferencing and comprehension as a whole.

Just as the effectiveness, proficiency level, and background knowledge of the listener can play a part in the listening strategies and processes that are employed, so, too, can task type. Recent research on language learning strategies has shown that what makes a successful L2 learner and user is more dependent on the learner's *choice* of strategies for a given task or situation (Cohen, 1996a, 1996b; Chamot & El-Dinary, 1999; Khaldieh, 2000, Vandergrift, 2007) rather than the actual *strategy*. In reviewing L2 listening strategy research, Chamot (2005) states that previous research has "confirmed that the good language learners are skilled at matching strategies to the task they were working on, whereas less successful language learners apparently do not have the metacognitive knowledge about task requirements needed to select appropriate strategies" (p. 116). This type of knowledge, described by Wenden (1991) as "the part of long-term memory that contains what learners know about learning" (p. 45), underlies learners' abilities to

"manage, direct, regulate, [and] guide their learning" (Wenden, 1998, p. 519). In fact, research on strategy use by effective and less effective listeners has found the use of metacognitive strategies and knowledge of task requirements to be particularly important for enhancing success (see O'Malley and Chamot, 1990; Vandergrift and Tafaghodari, 2010).

Although listening strategies researchers have uncovered a variety of factors that influence listening processes, they have primarily focused on the strategies that lead to listening successes rather than delving into processes of L2 listeners who face listening comprehension problems. In response to this gap, Goh (2000) identified real-time listening difficulties faced by a group of ESL learners and examined these within the three-phase model of comprehension proposed by Anderson (1985). A variety of data were collected in order to determine these difficulties: diaries from 40 learners, interviews from 17 of the 40 students, and think aloud reports from 23 of the 40 students. In all, she was able to identify ten problems which occurred during one of the three phases; five were related to the perceptual processing phase, three to parsing, and two to utilization. After comparing the problems of two groups of listeners at both the high and low proficiency level, Goh found that low-level learners had more problems with perceptual-level processing, although listeners in both groups expressed some of the same difficulties. For example, learners from both groups were found to have problems recognizing words they knew and quickly forgetting what they heard. Goh suggested "this was probably due to excessive demands from unfamiliar input on a limited processing capacity", and that "due to the recursive nature of comprehension processes,

mental representations from successful parsing were displaced by new input before they could be utilized" (p. 67). In contrast, lower-level listeners were found to have problems with attention, such as not hearing one part of a text because they spent too much time thinking about something they had just heard. Goh noted that both high- and low-level listeners paid close attention to problematic parts of the text in order to understand it, but that higher-level listeners used the metacognitive strategy of directed attention to "bring their attention back to the unfolding text and continue with listening" (p. 67). Goh concluded that "understanding learners' listening difficulties [is useful] as it pinpoints those places in cognitive processing where comprehension can break down" (p. 57).

In sum, research on processing input in L2 listening suggests that L2 listeners of high and low levels encounter different comprehension problems while progressing through the three phases, and learners differ in their use of listening strategies while processing aural input which impacts their understanding of the text as a whole. While these researchers suggest practical tips for helping listeners improve their listening ability, including listening strategy training and schema-building exercises, they also note the difficulty in training L2 listeners to "hold as much of the spoken text as possible in their limited capacity short-term memory, interpret the content before it is displaced by new input, and provide immediate listener response if that is required" (Goh, 2000, p. 71). Allowing learners repeated exposure to aural input may aid short-term memory retention and lead to gains in listening comprehension.

Repetition and Listening Comprehension

As a task variable, repetition is most often tied to strengthening information in memory and influencing overall comprehension. In their discussion of L2 listening comprehension theory and pedagogy, Nagle and Sanders (1986) note that "*rehearsal* (conscious and unconscious repetition) may strengthen an item in short-term memory", and that "for [second language acquisition (SLA)] theory, there is a consensus among researchers in memory that rehearsal is an important variable in fostering long-term retention as well" (p. 15). Ellis and Sinclair (1996) provide empirical support for this latter claim with their finding that language learners who were forced to rehearse, or repeat, Welsh utterances demonstrated "superior performance in…receptive skills in terms of learning to comprehend and translate [foreign language] words and phrases" (p. 243) when compared to learners who were prevented from articulating the same utterances. While it was unclear whether the advantage of repetition in this study lay in the articulation of the utterances, or the students' hearing of their own repetitions, the authors conclude that "short-term representation and rehearsal allows the eventual establishment of long-term sequence information for language" (p. 247).

In addition to its role in long-term retention, repetition and restatement of input as a task variable can allow learners more time to process information in the input as well as the relationships between syntactic forms (Hatch, 1983). Jensen and Vinther (2003) tested the hypothesis that after listening to video recorded dialogues the first time, eighty-four Danish second-year university students studying Spanish at the intermediate level would have the opportunity to comprehend meaning and store this information in their working

memory. Studying the effect of exact repetition and speech rate reduction on comprehension of dialogues seen in video recordings, the authors hypothesized that learners would try to extract meaning from an utterance during the first time listening, and that during the second time, learners would already have located "the problematic features in the stream of sound" (p. 380) which would help them focus on forms and therefore, help aid their detailed level of comprehension. They compared the results of two treatment groups, which listened to the video conversations a total of three times each although at different rates of speech (fast-slow-fast or fast-slow-slow), and a control group, on performance of an elicited imitation task. While there was no significant difference between the two treatment groups in terms of comprehension, students from both groups were found to comprehend the material significantly better than students in the control group. The authors thus concluded that repetition allowed students to first process meaning and then reformulate hypotheses about language form and meaning during the subsequent listening. O'Bryan and Hegelheimer (2009) reported similar findings in a case study investigating the strategies used by four ESL students when listening to repeated input. Based on verbal protocol data and the student's notes, the authors found that repetition allowed one low-intermediate student to "build up to more complex bottom-up processing strategies [during the second listening]…[by] using lexical and grammatical relationships to comprehend the input" (p. 26). This resulted in the student obtaining a more complete understanding of the input as opposed to the "disjointed summary" (p. 25) reported during the first time listening to the text.

Although no other studies have investigated the impact of repetition on comprehension

processes, additional studies have concluded that repetition has a positive effect on listening comprehension. Cervantes and Gainer (1992) investigated the effects of syntactic simplification and repetition of academic lectures on the listening comprehension of 76 university-level English as a foreign language (EFL) learners. The authors found that repetition resulted in significantly higher comprehension scores. Cabrera and Martinez (2001) found that making use of repetitions, comprehension checks, and gestures helped 60 EFL school children better follow a story told by their instructor.

While there is evidence that repetition leads to increased listening comprehension, little attention has been paid to the impact that the *structuring* of repetition has on listening processes and comprehension; this is despite the fact that studies often differ in the way repetition is offered to learners, yet the conclusions and recommendations for teachers remain the same. In the above literature, Jensen and Vinther (2003) repeat each sentence of the dialogue in turn. In contrast, O'Bryan and Hegelheimer (2009) provide repetition in a modified, whole-text format as the first time, the input was paused at pre-determined places in the text to allow for think-aloud data to be collected; the entire text was played in full the second time. Cervantes and Gainer (1992) offer exact repetition following pre-determined dictation segments, and Cabrera and Martinez (2001) offer repetition of words or phrases "taken from something said in the three previous utterances" (p. 285). Many researchers investigating computer-based listening materials allow learners to repeat whole texts, phrases, words, and even parts of words (see Pujolà, 2002; Hegelheimer and Tower, 2004; Grgurovic and Hegelheimer, 2007) due to the capabilities

offered by the medium. Yet because of these obvious differences in the way repetition is presented to students in all of these studies, the conclusions that repetition influences students' listening processes, as well as leads to gains in listening comprehension, is perhaps misleading to instructors hoping to improve their students' listening skills. This confusion is compounded by the fact that many academic listening textbooks (e.g. Dunkel, Pialorsi, and Kozyrev, 1996; Lebauer, 2000; Salehzadeh, 2006) suggest that teachers play an entire lecture or lecture excerpt in whole before it is repeated, a suggestion that is likely due to classroom constraints that prevent repetition at the sentence, or even word or syllable, level.

Although repetition may positively influence listening comprehension and comprehension processes, the issue of timing is one that deserves attention. Empirical work in repetition and listening comprehension does little to direct teachers in the way of providing repeated input that will reap the most benefits, e.g. positively influence listening comprehension processes and increase overall listening comprehension. Rather, it is a combination of theories from both psychologists and applied linguistics that, when combined with what is known about listening comprehension processes, can help teachers and researchers begin structuring repetition in a way that will be most beneficial to learners.

Enhancing attention and structure building through structured repetition

Referring back to Nagle and Sanders' (1986) model of listening comprehension processing in Figure 1, attention stimulates rehearsal—repetition of the input in the mind

as a way of retaining it in memory (Field, 2004)—in STS before the information is passed along to the executive for further processing; attention is also responsible for narrowing focus and monitoring, a metacognitive strategy observed in many studies on listening strategies (Bacon, 1992; Goh, 2002; O'Malley et al., 1989; Vandergrift, 1997). While the importance of attention in SLA has been widely documented (see Robinson's 2003 review), it can be compromised when a learner is unable to "sustain attention to a task and maintain the level of *effort* expended" (Robinson, 2003, p. 652, emphasis in original). Robinson notes that this failure can be caused by a prolonged time on task, as well as by the complexity of the task. This results in declining performance over time, such as "failure to correctly identify and interpret auditory input" (p. 652).

In order to ease the cognitive burden caused by both time on task and task complexity, psychologists Samuels and LaBerge (1983) propose dividing complex tasks into smaller processing units of shorter duration. In doing this, "the unskilled person can perform complex tasks by doing one subunit at a time" (p. 45). Rost (2002) argues that "chunking the input" is a mean of "achieving greater [listening] comprehension without altering a text" (p. 131). While the procedure of listening to chunked or segmented texts can be slow and difficult for the learner, through practice, "the attention demands for the subunits decrease, enabling the student to group the subunits into larger and larger chunks until the entire task can be handled as a single unit" (Samuels and LaBerge, 1983, p. 45). However, Samuels and LaBerge note the importance of "being careful that the energy demands of the subunits are less than our capacity limitations" (p. 45); in other words, if

the subunits are too long or too complex, they will continue to result in cognitive overload for the learner.

One suggestion for dividing texts in a way that will lessen the cognitive demands of the learner and enhance comprehension comes from the field of discourse studies. In his discussion of episodes as units of discourse analysis, van Dijk (1981) proposes that smaller processing units, termed 'episodes', can "have psychological relevance as units in a cognitive model of discourse processing" (p. 178) when organized around a particular theme or macroproposition. Van Dijk suggests that new episodes may be indicated through a number of grammatical signals, including pauses and hesitations in spoken discourse, paragraph indentations in written discourse, and time, place, and 'cast' change markers (e.g indefinite articles to introduce new individuals, and full noun phrases to reintroduce old one); "[it] goes without saying that such markers play an important role in a cognitive model for the strategies of discourse comprehension in which the language user has to derive a macroproposition [i.e. theme or main idea] from the propositions in the text" (p. 181).

Dividing texts based on macropropositions can assist learners in building a mental representation of the text as a whole. Drawing on his work with Kintsch (Kintsch and van Dijk, 1978; van Dijk and Kintsch, 1977, 1983), van Dijk (1981) notes that "the first sentence [in a text] is strategically used to derive a macroproposition. This macroproposition remains in Short Term Memory for the rest of the interpretation of the same episode. As soon as propositions are interpreted that no longer fit that

macroproposition, a new macroproposition is set up" (p. 191). In other words, an episode, which is thematically unified around a macroproposition, can help the learner appropriately 'chunk' information in a way that allows for "more structured representation in memory and especially better recall" (p. 191).

This idea is similar to Gernsbacher's (1990) Structure Building Framework where comprehenders use the initial sentences of paragraphs to "lay a foundation" (p. 5) for building a mental structure. Once this initial foundation is laid, subsequent input is "mapped onto a developing structure because the more coherent the incoming information is with the previous information, the more likely it is to activate similar memory cells" (p. 2). Following this process, memory cells will either enhance or suppress the activation of other cells depending on how relevant the information is for building subsequent structures. If the goal of comprehension is to "build a coherent mental representation…of the information" (p. 1), then enhancing structure building through the semantic chunking of information is especially important in lecture comprehension, which depends less on the meaning of individual sentences and more on their inter-relatedness and structure of the whole text (Dunkel and Davis, 1994).

In addition to building a mental representation of the text, semantic chunking may also help encourage listeners to focus on semantic cues in the input, thereby activating their semantic memory and encouraging more efficient input processing. Engaging one's semantic memory when processing linguistic input is advocated by Kutas and Federmeier (2000) who note that "[predicting content through the use of one's semantic memory]

allows for more efficient processing when the expectation is upheld" (p. 467). In addition, it also allows for "greater perceptual accuracy when the expected item is somehow degraded or garbled" (p. 467) as is sometimes the case with oral input. Therefore, engaging semantic memory through episodic chunks built around semantic cues can not only lead to more efficient input processing, but it can also allow listeners to overcome common difficulties (i.e. perceiving only garbled input) experienced when listening.

To sum up, providing repetition and dividing longer, more complex texts into shorter, thematically-based units can theoretically help L2 listeners create a structured representation of the text in memory, thereby enhancing recall. Repetition has been shown to give learners more time to process input for both meaning and form (Jensen and Vinther, 2003; O'Bryan and Hegelheimer, 2009), and both repetition and the division of tasks into smaller units have been shown to raise understandability (see Bremer and Simonot, 1986; O'Bryan and Hegelheimer, 2009). Bremer and Simonot (1986) echo van Dijk (1981) when noting that referential or semantic ambiguity requires additional effort on the part of the language learner; the authors provide support for this claim with a number of case studies documented in Bremer (1983) showing that making new topics salient and their content expectable increases learners' chances of understanding. Ellis (2003) argues that lexical, and even semantic, chunking increases the salience of input by allowing learners to activate associated "meaning representations" (p. 78); this, in turn, facilitates parsing, a crucial phase in the listening comprehension process. Therefore, providing repeated input through structured repetition could both enhance recall and

facilitate comprehension processes by raising understandability, making input salient, and assisting learners in making mental representations of the input through structure building.

By enhancing attention and structure building through structured repetition, it is conceivable that this type of task condition could help students overcome some of the listening comprehension problems identified by Goh (2000). For example, being able to maintain attention to the task at hand may allow learners to bypass some of the perceptual processing problems identified, including "miss[ing] the beginning of texts" and "concentrat[ing] too hard or [being] unable to concentrate" (p. 59). Overcoming these problems is crucial to learners getting the input into STS which can then be parsed and utilized. Likewise, overcoming parsing problems such as "quickly forget what is heard" (p. 59) can help learners keep information in STS long enough so that it can be appropriately linked to background knowledge or lived experience in LTS. Dividing tasks into macropropositions, which can assist learners in building a mental representation of the text as a whole, directly addresses the parsing problem of "unable to form mental representation" (p. 59). Something as simple as changing the way repeated input is presented may have a dramatic impact on the comprehension problems that learners experience.

Research Questions

Based on the review of previous literature, this study will investigate the following research questions:

1. Does task type (whole-text or segmented repetition) impact comprehension of the listening text?
2. Does task type impact the types of comprehension problems a learner experiences?

Based on van Dijk's (1981) theory that structuring texts using semantic episodes can help learners appropriately 'chunk' information in a way that allows for "more structured representation in memory and especially better recall" (p. 191), it is hypothesized that students in the segmented repetition condition will obtain higher listening comprehension scores than students in the whole-text condition. While students in both groups will experience comprehension problems, students in the whole-text condition are expected to encounter more problems attending to the task and forming a mental representation of the text compared to students in the segment group who are provided with more guidance in attending to the task and structure building through the design of the segmented task.

Summary

This chapter reviewed the literature that provides a background for this dissertation work. It reviewed models of listening comprehension and discussed ways in which input processing has been researched in the area of L2 listening comprehension. The impact of whole-text repetition on both comprehension and the processing of aural input was detailed, which led to a discussion of how attention and structure building might be enhanced through the provision of structured repetition. Finally, the research questions

were introduced. The next chapter describes the overall design of the study. Methods, including materials, measures, the procedure of data collection and data analysis, are explained in detail.

CHAPTER 3: METHODS

This chapter describes the methodology undertaken in this dissertation, which follows a mixed-methods embedded design to evaluate the impact of two repetition conditions, whole-text and segmented repetition, on students' overall listening comprehension, as well as on the comprehension problems they experienced. Quantitative and qualitative data collected from a total of 100 intermediate English language learners were used to answer the proposed questions. Characteristics of the participants, who included graduate and undergraduate students, are described based in part on information provided by the students themselves. Following that, a detailed account of the materials and data collection instructions is provided. A description of the procedures employed in both the quantitative and qualitative portions of the study is then provided, followed by an explanation of how the data was analyzed in order to address the proposed research questions.

Study Design

This mixed-methods study investigates the impact of two repetition conditions, whole-text and segmented repetition, on students' overall listening comprehension, as well as on the comprehension problems they experienced. Mixed methods research recognizes that "*both* quantitative and qualitative research [methods] are important and useful" (Johnson and Onwuegbuzie, 2004) and seeks to use both in either a single study or "among several studies in a program of inquiry" (Creswell, 2002, p. 210). Applied linguists and those in the area of educational measurement increasingly cite the strengths of mixed methods designs (see Creswell and Plano Clark, 2007; Ortega and Iberra-Shea, 2005). Using both

a mix of quantitative and qualitative data can be useful for investigating a complex issue like comprehension.

The design of this study follows an embedded design (Creswell and Plano Clark, 2007) where both quantitative and qualitative data were collected, "but one of the data types plays a supplemental role within the overall design" (p. 68). In an experimental research design, for example, qualitative data may be included "to examine the process of an intervention…or to follow up on results of an experiment" (p. 67), and both data types are used to answer different research questions within the study. In the present study, the quantitative data (i.e., written recalls) were collected from a group of participants that were semi-randomized following a method recommended by Mackey and Gass (2005), and discussed further in the "Procedures" section of this chapter. The quantitative portion of the study is identified in the center box in Figure 2, which shows a visual representation of the study design. Qualitative data is embedded within this experimental design before (i.e. pre-task survey), during (i.e. immediate retrospective verbal protocols) and after (i.e. post-task interviews) the intervention, as indicated by the small boxes on either side of, and also below, the middle, "quantitative" box.

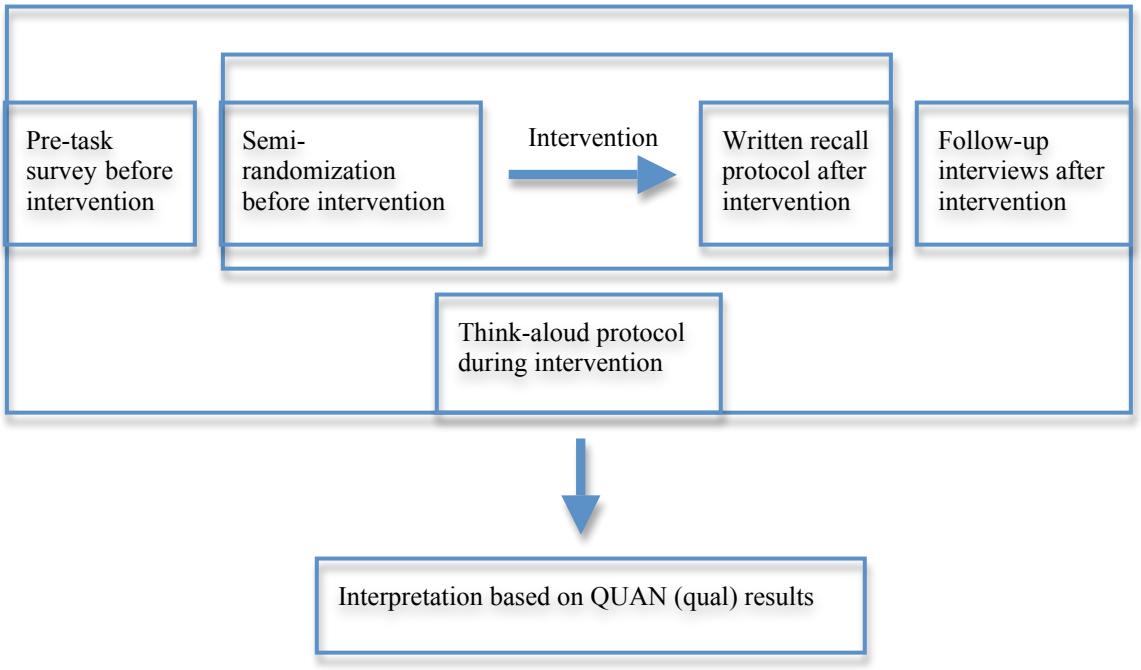


Figure 2. Research design model. Adapted from Creswell and Plano Clark (2007)

Participants

Participants in this study were 100 intermediate-level English language learners enrolled in an academic listening strategies class at a large state university in the U. S. Midwest. These participants are graduate and undergraduate students who have a high enough level of overall English proficiency to have been admitted to the University, but based on the results of an institutional English Placement Test (EPT), they were required to take an academic listening strategies course during their first year at the University. The number of participants was informed by both a pilot study and statistical power analysis. First, a pilot study with 12 participants—four each at the low, intermediate, and advanced levels of English proficiency—revealed that low-level listeners were found to encounter numerous problems with unknown vocabulary and, therefore, were kept from fully processing the input in the listening texts. In contrast, ESL students at the intermediate

level are ideal for studying impact of task type on comprehension processes and problems because their level of vocabulary can adequately allow them to engage in the processes necessary for comprehension. Second, a prospective power analysis was conducted prior to the research study for a two independent samples t-test in response to criticism in both L2 research (Crookes, 1991) and the behavioral sciences (Cohen, 1988) that it is essential, yet rarely done. Statistical power of a given test depends on three factors: alpha level, effect size, and sample size. Given an alpha level of $p = .05$, as well as a medium effect size ($d = 0.05$), a sample size of 128 would have been required to reach the desired power of 0.80 suggested by Cohen (1988). However, only 100 students agreed to participate in the research study, and therefore a power of 0.70 was obtained. This indicates that there is a 30% chance of accepting a false null hypothesis (i.e., committing a Type II error).

The participants were from 10 countries: China (83), Korea (4), Taiwan (1), Jordan (1), Spain (1), Colombia (1), Ghana (2), Malaysia (4), Japan (2), and Nepal (1). A complete participant overview can be seen in Table 1. Of these 100 students, 12 students (6 in each repetition condition) participated in the think-aloud portion of the study in order to shed light on the comprehension processes and problems experienced while interacting with texts in the two conditions. These 12 students were chosen based on a number of factors: the repetition condition into which they were placed, their nationality, gender, and overall listening proficiency as indicated by their scores on the listening portion of the institutional EPT. An overview of these students can be seen in Table 2.

Table 1

Participant Overview

Variable Categories	Participant Variables	Whole-text condition	Segmented condition
Sex	Males	30	26
	Females	20	24
Nationality	China	40	43
	Korea	1	3
	Malaysia	3	1
	Other	6	3
Student classification	Graduate	7	5
	Undergraduate	43	45
Time in the U.S.[a]	Mean	7.64	7.90
	Standard Deviation	6.83	6.60
EPT Listening scores[b]	Mean	12.52	12.15
	Standard Deviation	2.93	2.17

Note. [a] = Time measured in months; [b] = Maximum score on EPT Listening test = 30.

Table 2

Overview of students participating in the think-aloud portion of the study

Participant ID (W = whole-text; S = segmented)	Nationality	Sex	EPT Listening scores	EPT Listening Mean score[a]	EPT Listening SD
W003	Japan	M	14	13.3	1.50
W004	Malaysia	F	13		
W017	China	F	11		
W018	Ghana	M	13		
W035	China	F	13.2		
W050	Taiwan	M	15.6		
S020	Malaysia	F	11	12	1.41
S025	China	M	12		
S026	Japan	F	12		
S027	China	M	13		
S028	Korea	M	10		
S029	Korea	F	14		

Note. N = 12 (6 in each condition); [a] = Maximum number = 30.

Materials

Listening texts

As university students, an important aspect of their academic life is listening to academic lectures given by native or near-native English speakers to a general, graduate or undergraduate audience. Because understanding the lectures is a goal for these students, the listening texts used in this study were excerpts (about 3 minutes) from authentic academic lectures. Short excerpts, rather than whole lectures, were chosen for reasons of feasibility as well as to observe a period of unfatigued listening. As Ur (1984) notes, "in a long listening comprehension exercise a learner's grasp of the content is much better at the beginning and gets progressively worse as he goes on" (p. 19). While many factors may contribute to this, Ur notes that a major contributor is fatigue which causes the listener to "run[] out of the energy necessary to absorb and interpret strange sounds" (p. 20). Excerpts, or shorter versions of lectures, are often used in studies of listening comprehension (see Flowerdew, 1994, for a number of examples). In the present study students listened to excerpts from two, level-appropriate texts (see discussion below). The speakers were both non-fiction writers who had been invited to speak about the topics of their books at the Technology, Education and Design talks (TED talks) in Monterrey, California. Both speakers were male and had North American accents, although this was not intentional; rather, these two texts received the most consistent scores from the raters when rating overall difficulty (explained below). While the topics were general interest topics in the human or social sciences, the Jacobs (2007) lecture draws upon a number of cultural references (e.g., the Olive Garden restaurant, lying about someone's weight in order to preserve a relationship, etc.), whereas the Johnson

(2006) lecture is based on a historical event and therefore contains references to people and places from the time period. An overview of the texts can be seen in Table 3.

Table 3

Overview of listening texts

Speaker	Accent	Topic	Source
Male	North American	Cholera outbreak in 17th century London	Invited lecture by Steven Johnson (2006) presented at the Technology, Education and Design (TED) talks
Male	North American	Researching possible writing topics for the lecturer's magazine articles	TED talk by A. J. Jacobs (2007)

A number of steps were taken in the pilot phase of the study to ensure that the listening texts were of an appropriate level and topic for students at this level. First, materials used in the ESL listening class from which the participants were recruited were examined for examples of typical lecture speed, lexical difficulty and authenticity. With these factors in mind, four different lecture excerpts from the social sciences were chosen as possible texts for students at this level. Next, each text was analyzed by calculating the speech rate, type-token ratio, and studying the lexicon by entering each text into the Web Vocabprofile developed by Cobb (2001); this profiler allows a user to submit written text to an online vocabulary 'sorter' which "breaks texts down by word frequencies [based on Laufer and Nation's Lexical Frequency Profiler, 1995]…and divide the words of texts into first and second thousand levels, academic words, and the remainder or 'offlist'" (Cobb, 2006, VocabProfile Home). This preliminary analysis of content, speech rate, discourse and lexical complexity resulted in an initial ranking of overall text difficulty.

Following this preliminary analysis, a rating rubric was designed so that four outside raters who were experienced language teachers could also rate each text for overall difficulty (see Appendix A). Definitions based on the work of Brown (1995) and Buck (2001) regarding factors that increase the listening difficulty of academic speech, including content difficulty, syntactic complexity, discourse structure and lexical complexity, as well as additional factors affecting the complexity of all aural input, including speech rate, accent, and prosodic features (i.e. rhythm, stress, intonation), were included in the rubric to be used in assessing overall difficulty level of each of the nine texts. Intra-class correlation coefficients (ICCs), which represent the ratio of the variability between raters to the variability within raters, were computed to check interrater reliability. The ICC is useful because it can be used when there are more than two raters. The range of the ICC values run from 0 to 1; the average measures ICC for these four raters was 0.83, meaning that there was little variation between the raters compared to variation among rater means (see Shrout and Fleiss, 1979). The two texts used in the present study were ultimately chosen by reconciling both the initial ranking of overall text difficulty and the raters' rankings, i.e. texts used in the study are those that received similar scores from both ranking criteria.

Texts were prepared slightly differently for the classroom and think-aloud portions of the study. In listening textbooks, texts presented in the "whole" condition are played through once, and then again. When qualitative, process data have been collected in past research, texts presented in the "whole" condition are not edited, but are paused to allow for think-

alouds to be collected. Vandergrift (2003) states that these pauses occur at "natural discourse boundaries" (p. 473); no additional detail is given that elaborates on how these natural boundaries are identified. In the present study, guidelines for texts presented in the segmented condition draw on van Dijk's (1981) notion of the episode. These are "characterized as coherent sequences of sentences of a discourse, linguistically marked for beginning and/or end, and further defined in terms of some kind of 'thematic unity'" (p. 177).

In order to determine an appropriate placement of pauses in the whole-text repetition condition (i.e. the 'natural discourse boundaries') before process data was collected, as well as the division of texts in the segmented repetition condition, all texts used for qualitative data collection were divided into semantic units using van Dijk's (1981) method of analyzing a text based on the notion of the macroproposition, intuitively known as the 'theme' or 'gist'. Van Dijk states that a new macroproposition should be formed when the "cast of participants, time, place, circumstances, and (global) event or action seems to change" (p. 191). Because the division of text into units is based on a set of semantic criteria, the length of each macroproposition may differ; for example, the introductory episode may last only 20 seconds, while the next semantic episode, organized around a sub-theme of the lecture, may last for forty seconds. Specific examples of the texts and semantic episodes used in the study can be seen when looking at the weighted rubrics in Appendix B.

Using the examples in Appendix B of semantically segmenting the lecture as a guide, students listening to this text who were in the whole-text group for qualitative data analysis were presented with segment 1, and then the text was paused to allow for the think-aloud to be collected. Following this, students listened to segment 2, and then the text was paused for the think-aloud. This procedure was followed for the entire text. Again following the procedure used in past research (see Vandergrift, 2003), once the student listened to the whole text once, they were told that they would listen to the text again in full. They were asked to tell the researcher if they had additional clarifications or insight to add, and the text was paused at that time for the think-aloud. Students in the whole-text group for quantitative data analysis listened to the whole text once, and then again. Students in the segmented repetition group for both qualitative and quantitative data collection were presented with one segment at a time, however each segment was repeated in turn. In other words, at no time did the students in the segmented repetition group listen to the entire text in full.

Digital recording equipment

The verbal reports were recorded using a computer equipped with digital audio recording software and an external microphone.

Qualitative Analysis Software

A software program designed to facilitate qualitative data analysis, Transana (Woods and Fassnacht, 2007), was used to facilitate the coding process and create a seamless link

between the coding scheme and particular points on the digitized file containing each verbal report.

Recall Protocol Rubrics

Rubrics were created during the pilot phase to rate student recalls and later revised. The method of scoring the recall protocol has varied in the literature, but the basic procedure is similar: break each lecture excerpt into units and compare the ideas present in students' recall protocols to the units from the lecture excerpt. However, researchers have varied in their approaches to breaking the original text, whether oral or written, into units. Both Meyer (1975) and Berhnardt (1983) "slash[] prose into meaningful segments" (p. 29), while Carrell (1985) divides her texts into 'idea units' which, Alderson (2000) notes, are "somewhat difficult to define" (p. 230). Johnson (1970) devised an "objective method…for dividing complex verbal materials into linguistic subunits possessing psychological significance" (p. 12). He suggests breaking verbal passages into pausal units based on the notion that "locations in the story which [are] psychologically acceptable for pausing [are] hypothesized to be one of the functional boundaries used in encoding and decoding [input]" (p. 13). Due to the important role of pauses in processing spoken input, lecture excerpts were first broken into pausal units as the first step in creating a scoring rubric for students' recall protocols.

After dividing the two lectures used in the study into "pause acceptability units" (Johnson, 1970, p. 13) in which pausing may have been used to "catch a breath, give emphasis to the story, or to enhance meaning" (p. 13), the texts were further analyzed to omit the pausal units containing redundancies, speech infelicities, or phrases lacking in

content related to the topic (e. g. “And I’ll do—I’ll define that later on”) because it would be unrealistic to expect participants to recall these types of pausal units verbatim. Finally, the text was further analyzed by two composition teachers and reduced to the pausal units which provided either a given main idea or a given detail only once. The final number of pausal units for each lecture was 56 for the Jacobs (2007) lecture and 43 for the Johnson (2006) lecture.

Once the final collection of pausal units remained for each text, the Johnson (1970) method was used to weight each pausal unit based on its importance to the overall text. Each pausal unit was weighted on a scale from one to four, depending on its importance to the meaning of the passage, with one being the least important and four being the most important. Two independent raters with backgrounds in applied linguistics were asked to weight each pausal unit using the four-point scale (see Appendix C for rater instructions). Similar to Johnson’s (1970) study, raters’ scores were correlated in order to calculate interrater reliability. Mackey and Gass (2005) suggest using Spearman rho for correlating ordinal data; the resulting correlation coefficient was $r = 0.73$ for the Jacobs (2007) lecture, and $r = 0.74$ for the Johnson (2006) lecture. Both of these correlations were significant at the $p = 0.01$ level. Disagreements were settled by a third rater, and a fourth rater was brought in for the rare cases when the other three did not agree. The results were rubrics to be used when scoring each subjects’ recalls (see rubrics in Appendix B).

Measures

Pre-task survey

A pre-task survey was given to all students in which demographic information and information on the confounding variable topic familiarity was collected. For the former, information about classification (graduate/undergraduate), nationality, native language, age, sex, length of time in the U.S. and area of interest/major was collected. For the latter, students were given a list of twenty topics, in which eight areas related to the topics of the lectures were included (e.g. Disease Outbreaks, Magazine writing), to rate on a Likert scale of 1-5 with 1 being "no knowledge of the topic" and 5 being "expert on the topic". A copy of the pre-task survey can be seen in Appendix D.

Written recall protocol

To assess listening comprehension, a recall protocol was used in which, immediately after listening to the lecture, subjects were asked to write as much as they felt they understood from the lecture (see Appendix E for the recall protocol prompt). The recall procedure has been frequently used in reading comprehension research (see Bernhardt, 1983; Carrell, 1985), and is becoming a common measure of comprehension in listening research (see Graham, Santos, and Vanderplank, 2008; Hahn, 1999; Schmidt-Rinehart, 1994). The recall protocol is thought to "circumvent the pitfalls of traditional test design" (Bernhardt, 1983, p. 28), and Alderson (2000) states that this technique is "often held to provide a purer measure of comprehension, since test questions do not intervene between the reader and the text" (p. 230).

While the recall protocol is easy to administer, there are a number of issues that must be taken into consideration when designing the task. First, Alderson (2000) states that the recall should be given "in the first language, otherwise it becomes a test of writing as well as [the language skill being tested]" (p. 230), although he notes that in many studies of English as Foreign Language (EFL) readers are asked to write the recall in their L2. In the present study, students were given the option to write either in their L1 or L2; all students in the present study chose to write in their L2 (English). Chang (2006) also found indication "that the requirement of memory in the recall task hinders test-takers' ability to demonstrate fully their comprehension of the reading passage" (p. 520), although Alderson (2000) states that "if the task follows immediately on the [input], this need not be the case" (p. 232). Note-taking during the task, in addition to adding to the authenticity of the task, could also help combat the issue of memory. Finally, while the task may be administered orally, Riley and Lee (1996) found that significantly more main ideas were recalled when learners were asked to provide a written recall. The format of the recall in the present study was a written recall to be completed, in a language of their choice, using pencil and paper.

Verbal protocol

Verbal or "think-aloud" protocols were used with students in the think-aloud portion of the study to elicit on-line or real-time comprehension processes used while listening (see Ericcson and Simon, 1984). With this method, individual learners are asked to voice their thoughts while working through a language task. Such introspective methods "attempt to overcome a principal problem in psycholinguistics: the desire to describe a learner's knowledge about a language based on incomplete evidence stemming from learner

production" (Gass, 2001, p. 221). In collecting verbal protocol data, many times an interviewer sits with a student while he/she completes the tasks and asks questions such as "What are you thinking?" and "Why did you decide to do that?" Gass (2001) notes that asking the right questions during a verbal protocol procedure, in addition to training participants in the procedure beforehand and setting up an appropriate structure for conducting the think-aloud sessions, can help to avoid some of the common pitfalls of the think-aloud method. Some researchers have questioned the validity of learners' verbal reports, questioning whether learners were actually "giving evidence of an inner process…or whether they were saying things to [the researcher] that might be different from what they'd say to someone else" (Smagorinsky, 2001, p. 234), or whether these reports might actually alter the cognitive processes learners employ (see Ericsson and Simon's (1984) discussion of and response to these criticisms). While introspection has not always been accepted as "a valid tool for gathering information about knowledge of language…it is now being used once again with some frequency and with increased confidence" (Gass, 2001, pp. 226-227). In the present study, students were trained in thinking out loud with a sample text before listening being recorded, were able to think aloud in either their native or second language, and the think-aloud was recorded and analyzed afterwards for evidence of learning strategies students used and comprehension problems encountered while listening.

Semi-structured interviews

Short, semi-structured interviews took place after learners in the think-aloud portion of the study listened to each text. The purpose was to elicit the learners' own evaluations of

what they understood, what they did not understand, what strategies or processes they felt they used while making meaning of the text, and what impact they felt the structure of the task had on their comprehension and comprehension processes. This information helps support findings gained from the recall protocols and verbal protocol data.

Procedure

The present study uses a mix of both quantitative and qualitative methods. Therefore, the procedure differed for students in the classroom (quantitative) and think-aloud (qualitative) portions of the study.

Classroom study

Because the participants were enrolled in an ESL listening class at the time of data collection, the classroom portion of the study was carried out with intact classes, i.e. up to 20 students at once. While Mackey and Gass (2005) acknowledge that this is not typical of experimental research, they note that using in-tact classes is not only more feasible for students with varied academic and work schedules, but "may have the advantage of enhancing the face validity of certain types of classroom research" (p. 143). The authors note that while randomization of individuals may not be feasible, one option is to "use a semi-randomization procedure by arbitrarily assigning classes to one treatment or another" (p. 143). This semi-randomization procedure was followed in the present study.

Data collection took place in a computer lab. Upon beginning the research study, informed consent was obtained from all participants; following this, students completed the pre-task questionnaire. They were then told the purpose of the research study: to find out whether one type of repetition condition leads to higher listening comprehension than another. They were informed of which condition they would be experiencing, whether whole-text or segmented, and the design of the texts was described (see section *Listening Texts* above). Next, each student was given paper for taking notes during the lectures and for writing their immediate recalls. Before listening to each lecture, students were told the topic of the lecture. Students were given this opportunity to activate schema as it more closely resembles an authentic academic listening task where students typically have knowledge of the topic and, often, specific content before listening to a lecture. Finally, after listening to each lecture excerpt, students were asked to write down everything they felt they understood from the lecture using either their native or second language (i.e., English); these instructions were provided both orally and written at the top of the paper on which the students' recalls were to be written. Students were given 15 minutes to write their recalls; most students completed their recalls before the 15-minute time limit. In all, students in the classroom portion of the study took no more than 50 minutes (a typical class period) to complete the study.

Because there were two lectures that students were listening to and recalling, a counterbalanced design in which the ordering of the test items (i.e., the presentation of the listening texts) is different for different participants was adopted. For example, in

two, in-tact classes that were both listening to texts in the segmented repetition condition, one class listened first to the cholera lecture (Johnson, 2006), completed the recall, and then listened to the writing lecture (Jacobs, 2007) and completed the recall (see Table 3 for an overview of the texts used in the study); the reverse order was followed for the second class. Counterbalancing is one way to compensate for a lack of comparability due to non-random assignment, as is the case in this study (Mackey and Gass, 2005).

Think-aloud study

Students participating in the think-aloud portion of the study indicated their interest in doing so on their pre-task surveys completed during the quantitative data collection that took place in their listening class. The think-aloud session was scheduled as close to the classroom data collection as possible, although because the session time depended on students' individual schedules, the time between the classroom and think-aloud data collection session varied anywhere from one to three weeks apart. Students in the think-aloud portion of the study were first told that the purpose of the session was to gain information about their listening comprehension processes and problems and that they would be talking out loud about anything they were thinking while listening. This information was read from a standard script in order to ensure all participants received the same information. Next, students completed a training session (using an actual lecture excerpt) prior to data collection so that they would have a thorough understanding of the think-aloud procedure and ample opportunity to practice. Students were given guidance about how to "think out loud" during the training and were allowed to practice until they felt comfortable with the process. All data collection sessions were conducted on an

individual basis and audio-recorded for later verbatim transcription and coding. Sessions lasted approximately 1 hour.

Think-aloud data were recorded for the two different texts. In the whole-text repetition condition, students first listened to one segment before the text was paused, giving them the opportunity to tell the researcher what they were thinking as they listened to that segment and voice any problems they had understanding the text. If the student was unsure of what to say, the researcher asked questions such as "What are you thinking?" and "How did you figure that out?" Following this think-aloud, the researcher immediately played the next segment; students were instructed to think out loud following that segment, and so on. After listening to each segment once, the entire text was played in full. Students were asked to pause the lecture if they had something to add or change from their original think-aloud or about problems they were experiencing. In the segmented repetition condition, students again listened to one segment before it was paused for the think-aloud. However, once the student finished his/her think-aloud the segment was again repeated and followed by another pause for the student to think out loud. This process was repeated for each subsequent segment until each had been heard twice. Similar for students in the classroom portion of the study, students did not approach the text "cold" since they were reminded of the topic of the lecture before listening to each excerpt. Also similarly, students were allowed to take notes while listening to each text in order to more closely replicate an authentic academic listening task.

Data Analysis

Research Question 1

To address research question 1 regarding whether task type (whole-text, segmented) impacts comprehension of the listening texts, the participants' written recall protocols were examined for evidence of comprehension. Each protocol was scored according to the presence or absence of the main idea of each pausal unit represented in the scoring template (see section on Recall Protocol Rubrics above for complete description). Although pausal units were very short in length (see Appendix B for the recall protocol rubrics), students would sometimes recall just a portion of a pausal unit. It was determined that if that part was *not* the main idea of the pausal unit, they would get a score of 0. An example of a recall where the student recalled a portion of the pausal unit that was not the main idea (and therefore received a score of 0) can be seen in Table 4. One other common issue encountered in student recalls was the misspellings of words and incorrect placement of word boundaries in their written recalls. In this case, raters had to use their knowledge of ESL student phonological errors in order to determine whether they *understood* what the word was even if they could not identify this word once it was heard due to their underdeveloped listening vocabulary. An example of a word boundary error that resulted in a full score can be seen in Table 5. In addition, students were sometimes too general in their recalls. In order to determine whether the student understood what was in each pausal unit, raters were asked to assess the recall based on what was written by the student vs. what was implied.

Table 4

Example and justification of a partial pausal unit recall receiving a score of 0

Student Recall	Pausal Unit from Rubric/Lecture	Justification
"All around me is Italian"	"I'm Jewish in the same way the Olive Garden is Italian."	Student is not getting the main point of this idea unit where the author states he is Jewish.

Table 5

Example and justification of a word boundary error receiving a full score

Student Recall	Pausal Unit from the Rubric/Lecture	Justification
"His life is serious experiment."	"I see my life as a series of experiments."	Student misses the "series of" part, mistaking this for "serious," but does understand that the lecturer's life is some kind of experiment.

Student recalls were all coded by the researcher. Cohen's kappa was used to calculate intrarater reliability, which resulted in a value of 0.91. To check inter-coder reliability, a trained second rater with expertise in applied linguistics coded 25% of the data independently. Mackey and Gass (2005) suggest that considering the nature of the coding scheme is helpful in determining how much of the data should be coded by a second rater; the more objective and less inferential the coding scheme, the greater the chance that confidence in rater reliability can be established "with as little as 10% of the data" (p. 243). The recalls coded by the second rater were chosen randomly, with an equal number of recalls coming from students in the whole-text and segmented conditions. The second rater was informed that the purpose of the study was to determine whether two repetition tasks would result in different levels of student comprehension; the hypothesis that students in the segmented condition would recall more than students in the whole-text

condition was not shared in an attempt to maintain objectivity. Next, the second rater was trained by going through two sample texts, discussing common problems, and looking at a number of examples (see Appendix F for recall coding instructions). In order to maintain a high level of coder reliability, the second rater was asked to rate 5 texts at a time; discrepancies were discussed an resolved through adjudication. Cohen's kappa was used to calculate final inter-coder reliability, which resulted in a value of 0.83. Anything above 0.81 is deemed "excellent" (Mackey and Gass, 2005, p. 195).

After rating each written recall for comprehension, the total, weighted score for each item recalled was tallied for each text; next, these total weighted scores were added together so that each student ended up with one score reflecting the sum of their weighted scores for both texts. Next, a two-tailed, two independent samples t-test was conducted to compare the mean values of students' weighted scores in the whole-text and segmented repetition conditions. Although there is theoretical support provided in the literature review that would allow for use of a one-tailed t-test, because this issue has not yet been investigated, the two-tailed t-test was used in order to allow for rejection of the null hypothesis in either direction. In this t-test, listening comprehension scores were identified as the dependent variables that were affected by the factor "condition". The alpha value was set at $\alpha_2 = .05$, a common alpha level used in quantitative applied linguistics research (Hatch and Lazaraton, 1991). The dependent variable "listening comprehension score" was also assessed for effect size using Cohen's *d*-test in order to facilitate any subsequent meta-analyses incorporating these findings, as well as to provide support for the reliability of the findings. In order to gain greater understanding of how the task, and text, impacted

comprehension, a two-tailed, two independent samples t-test was also conducted to compare the mean values of students' weighted scores in the whole-text and segmented repetition conditions for each text. Cohen's *d* was again calculated.

The variable "topic familiarity" was investigated as a possible confounder using data from students' pre-task surveys. Participant responses on each of the eight areas related to the topics of the lectures (4 topic areas for each lecture) were correlated with the listening comprehension scores from each text using the Spearman rho correlation. This statistic is commonly used with interval data (the listening comprehension scores) and ordinal (Likert scale) data (Garson, 2008).

Research Question 2

A qualitative approach was taken to answer research question 2, which is concerned with the comprehension problems the students experience. First, all think-aloud and interview data were transcribed verbatim. Next, 25% of the think-aloud transcripts (6 of the total 24 transcripts) were coded by the researcher for comprehension problems using an open-coding approach. When appropriate, categories were named using the problems identified in Goh's (2000) research in order to facilitate discussion and comparison between the two studies by future researchers. Faerch and Kasper (1986) define comprehension as taking place "when input [from the text] and [a student's] knowledge are matched against each other" (p. 264); when there are "gaps in either input or knowledge, the [student] activates *inferencing* procedures, i.e. qualified guesses made on the basis of any information available" (p. 265). Therefore, a comprehension problem was defined as a time when input and knowledge were mismatched, whether due to an incorrect inference or a lack of

background knowledge about the topic. The end result of this coding process was a taxonomy to be used when coding the remainder of the think-aloud data (see Appendix G for a comprehensive list of listening comprehension problems, definitions, and verbatim examples).

Once the initial six transcripts were coded by the investigator, a second rater, who was a doctoral student in applied linguistics with experience teaching ESL students, was trained by going through the sample coding taxonomy with the researcher, looking at sample student transcripts that had been coded for comprehension problems, and then coding two transcripts together with the researcher (see Appendix H for second coder's instructions for coding comprehension problems). Discrepancies were discussed and, if necessary, the taxonomy was modified in order to more accurately identify the problems seen in students' think-alouds. Following Mackey and Gass's (2005) suggestion, interrater reliability was calculated not for the training data, but instead for the second set of eight transcripts that were double coded. Interrater reliability was calculated using a simple percentage which yielded an interrater reliability coefficient of 0.75. Intrarater reliability resulted in a reliability coefficient of 0.80. Simple percentages are often used in research investigating issues such as listening comprehension problems and strategies (see Goh, 2002; Young, 1997). Discrepancies were dealt with through adjudication. Following this joint coding session, the remaining ten transcripts were coded by the researcher. As different problems were coded, the frequency with which they occurred was also noted. If a student mentioned a problem with the same part of the text more than once, it was not counted as a new report; however, if a problem reoccurred throughout the text, e.g. if a

student continued to miss information at the beginning of each segment, then this was counted as a new report each time it occurred. Because students in the segmented condition were given more planned opportunities to talk during the think-aloud procedure (see *Think-aloud study* above), the total number of comprehension problems reported for each student was "normed" in order to account for the amount of time students talked. The frequency of comprehension problems was then reported as problems per minute of time the participant talked (talk time).

Summary

This chapter outlined the mixed-methods research methodology employed in the dissertation. It first explained how the embedded design (Creswell and Plano Clark, 2007) was applied to the context of the study in order to assess the impact of whole-text and segmented repetition on listening comprehension and comprehension problems. Next, information about the participants was provided. Following the description of the materials and measures used in the study, a detailed description of the classroom (quantitative) and verbal protocol (qualitative) procedures was presented. Finally, the specifics of how the data was analyzed in order to address the proposed research questions was addressed. The next chapter details the findings of the study.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter presents and discusses the results concerning both the impact of task type on listening comprehension and comprehension problems. First, the results for the first research question are introduced. They are based on the quantitative, written recall data, as well as on the topic familiarity information collected from the pre-task survey. The findings show that topic familiarity did not affect comprehension, and that task type had no statistically-significant difference on written recall (i.e., comprehension) scores. A number of issues were raised that might have influenced the results of this first research question; (1) the issue of whether segmenting a text is more appropriate for certain types of lectures than others, (2) the role of memory (and forgetting), and (3) the ability of the task to hold students' attention.

Following this, the results of the second research question are presented and discussed. These results are based on verbal protocol and post-task interview data. The findings show that students in both repetition conditions experienced problems during the three phases of comprehension, although students in the segmented repetition condition encountered fewer problems related to their ability to form a mental representation from words heard. This problem was directly addressed through task design, and supports the hypothesis set forth in chapter 2. Finally, evidence from the verbal protocols support the hypothesis that segmenting a text would enhance attention to the task.

Research Question 1: Impact of task type on listening comprehension

The first research question investigated whether task type (whole-text or segmented repetition) would impact listening comprehension when listening to academic lectures. Listening comprehension was measured using a written recall protocol that was scored using the Johnson (1970) procedure. The Johnson scoring procedure treats the pausal units in each recall protocol as individual items that are then summed to arrive at a total score. Before looking at descriptive and inferential statistics based on these scores, internal consistency estimates are needed "in order to determine whether the total scores are meaningful and appropriate" (Deville and Chalhoub-Deville, 1993, p. 125). However, Deville and Chalhoub-Deville (1993) noted that calculating internal consistency with weighted scores may result in an artificial score-spread and, therefore, an artificially high level of internal consistency. Therefore, the Cronbach's alpha reliability estimate was obtained for the recall protocol using recalls that had been scored both dichotomously and using the weighted rubric. The resulting values were 0.84 (Jacobs (2007) lecture, N = 56 items) and 0.84 (Johnson (2006) lecture, N = 43 items) for the dichotomously-scored data, and values of 0.84 (Jacobs (2007) lecture) and 0.83 (Johnson (2006) lecture) for data scored using the weighted rubric. These estimates indicate that the responses to the items are consistent and provide additional information about the quality of the recall protocol measure. In addition, they support Deville and Chalhoub-Deville's (1993) finding that "there is essentially no difference in subjects' relative total scores whether the recalls are scored dichotomously or are weighted" (p. 126).

Impact of task type on written recall scores

Descriptive statistics in Table 6 reveal that students' recall scores in the whole-text condition ranged from a total of 125 (W008) to 0 (W009). Student W009, who earned a score of 0 for both lectures, did participate in the recall task; however, none of the ideas the student wrote were consistent with those presented in the lectures.

Table 6

Recall scores for students in the whole-text repetition condition

	Text		
Participant ID	Jacobs (2007)	Johnson (2006)	Total recall score[a]
W001	35	10	45
W002	45	25	70
W003	27	27	54
W004	32	53	85
W005	20	27	47
W006	33	50	83
W007	47	35	82
W008	66	59	125
W009	0	0	0
W010	38	29	67
W011	10	12	22
W012	30	32	62
W013	19	20	39
W014	37	16	53
W015	28	22	50
W016	13	18	31
W017	33	3	36
W018	26	29	55
W019	23	13	36
W020	7	3	10
W021	14	12	26
W022	42	19	61
W023	9	10	19
W024	24	10	34
W025	68	53	121
W026	62	53	115
W027	49	37	86
W028	73	48	121
W029	8	6	14

Table 6 (Continued)

	Text		
Participant ID	Jacobs (2007)	Johnson (2006)	Total recall score[a]
W030	64	34	98
W031	40	43	83
W032	55	38	93
W033	31	20	51
W034	0	6	6
W035	11	11	22
W036	35	13	48
W037	19	22	41
W038	25	20	45
W039	29	16	45
W040	36	22	58
W041	64	57	121
W042	3	17	20
W043	4	4	8
W044	13	13	26
W045	34	39	73
W046	11	10	21
W047	35	25	60
W048	3	7	10
W049	15	18	33
W050	27	10	37
Mean	29.44	23.52	52.96
Standard Deviation	19.15	15.85	33.19

[a] Maximum score using the weighted rubric= 252. Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

Recall scores for students in the segmented repetition condition can be seen in Table 7. In slight contrast to students in the whole-text condition, the range of scores for students in the segmented condition is 120, with the highest score being 127 (S003), and the lowest being 7 (S012 and S044). While students from the segmented repetition condition received slightly higher mean scores on the recalls for the Jacobs (2007) lecture (30 vs.

29.44 for students in the whole-text condition), they received lower scores on the Johnson (2006) lecture (19.96 vs. 23.52) and with the two texts combined (49.96 vs. 52.96).

Table 7

Recall scores for students in the segmented repetition condition

	Text		
Participant ID	Jacobs (2007)	Johnson (2006)	Total recall score [a]
S001	58	31	89
S002	20	24	44
S003	56	71	127
S004	21	3	24
S005	12	18	30
S006	65	57	122
S007	7	7	14
S008	50	48	98
S009	11	13	24
S010	19	7	26
S011	22	16	38
S012	4	3	7
S013	76	30	106
S014	38	10	48
S015	63	11	74
S016	40	28	68
S017	49	38	87
S018	26	24	50
S019	24	16	40
S020	8	44	52
S021	28	8	36
S022	31	13	44
S023	20	13	33
S024	30	34	64
S025	72	13	85
S026	40	30	70
S027	49	27	76
S028	9	16	25
S029	21	18	39
S030	29	20	49
S031	55	38	93
S032	19	25	44
S033	9	0	9
S034	44	32	76
S035	20	7	27
S036	19	3	22

Table 7 (Continued)

	Text		
Participant ID	Jacobs (2007)	Johnson (2006)	Total recall score [a]
S037	53	13	66
S038	16	6	22
S039	42	35	77
S040	16	3	19
S041	19	21	40
S042	27	16	43
S043	28	29	57
S044	4	3	7
S045	10	13	23
S046	17	7	24
S047	14	4	18
S048	38	13	51
S049	33	15	48
S050	19	24	43
Mean	30	19.96	49.96
Standard Deviation	18.60	14.89	29.51

[a] Maximum score using the weighted rubric= 252. Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

When looking at students' recall scores, it is important to note that these scores reflect those obtained through use of the weighted scoring rubric, as detailed in Appendix B. It is clear that for the two lectures combined, students in both the whole-text and segmented conditions recalled pausal units that totaled only about 20% of the 252 points possible (the exact numbers are 21% and 20%, respectively). This number seems alarmingly low. Therefore, 2 native English speakers were asked to complete the same task; their recall scores are seen in Table 8.

Table 8

Recall scores for Native English speakers

Participant ID	Text		Total recall score [a]
	Jacobs (2007)	Johnson (2006)	
Native 001	85	79	164
Native 002	93	74	167
Mean	89	76.5	165.5
Standard Deviation	5.66	3.54	9.19

[a] Maximum score using the weighted rubric= 252. Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

As shown in Table 8, the two native English speakers recalled an average of 165.5, or 68%, of the 252 points possible. Although this is nearly three times the amount of information recalled by the ESL students, the native English speakers were unable to recall 100%--or close to 100%--of the ideas in the lectures. Therefore, having this information on native speaker comprehension, as well as their comprehension scores on this particular recall task, helps to keep the participants' recall scores in perspective.

While looking at the overall listening comprehension scores of students is helpful in understanding how much was recalled, it does little to describe exactly *what* was recalled. Looking at each group's frequency counts for the number of weighted pausal units recalled can provide this information. As a reminder, the pausal units of each text were weighted on a scale of 1-4 based on the importance of each pausal unit to the meaning of the passage (1 = least important, 4 = most important; see Chapter 3 for more details). Having these ratings makes it possible to look at the kind of information recalled most

frequently by students from each condition. Table 9 shows that, descriptively, there is very little difference in the total number of pausal units recalled at each weight for students in the whole-text and segmented conditions. Consistent with the information from Tables 6 and 7, students in the whole-text group recalled more pausal units from almost every weighted category. However, students in the segmented condition recalled slightly more pausal units with a weight of 3 (3 = very important) than students in the whole-text condition. This could have been aided by the task design, which divided each segment based on the important topics and sub-topics discussed in the lecture. Students from both conditions recalled more of the important ideas (weights 3 and 4) in the text than the less-important ideas (weights 1 and 2), and for students in both groups, there is a gradual increase in the average number of less-important to more-important weighted pausal units recalled. This helps the reader to see that ESL students at the intermediate level are able to identify, pay attention to, and recall more of the important ideas from listening texts than those that are not important.

Table 9

Total number[a] of pausal units recalled by weight

Condition	Weight of 1	Weight of 2	Weight of 3	Weight of 4
Whole-text	24.76	50.48	113.85	146.11
Segmented	17.62	47.14	117.18	123.89

Note. N = 100 (50 in each condition). [a]Due to the differing number of pausal units weighted 1-4 throughout the two texts, the total number of each weighted pausal units recalled for each group was normalized to 10. This allows a comparison to be made across the four weightings. Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

While the task design may have helped students in the segmented repetition condition to recall slightly more pausal units with a weight of 3 than students in the whole-text condition, the information in Table 10 shows that there was no significant difference in the written recall scores between students in the whole-text and segmented conditions ($p = 0.63$). The effect size ($d = 0.13$) fell below the small value suggested by Cohen (1988), indicating that a difference in comprehension scores is not present in the sample.

Table 10

Impact of task type on recall scores

Condition	M	SD	t	p	d
Whole-text	52.96	33.19	0.48	0.63	0.13
Segmented	49.96	29.51			

Note. N = 100 (50 in each condition). Maximum score using the weighted rubric= 252.

Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

Impact of task type on recall scores of each text

When taking a closer look at how students performed on each listening text individually, Table 11 shows that the results are consistent with the overall finding that there is no statistically significant difference in students' comprehension scores due to the condition in which they were placed. In other words, the different lecture did not change the impact that condition had on students' listening comprehension scores. One interesting finding to note, however, is the small effect size indicated for the Johnson (2006) lecture ($d = 0.23$). For this lecture, students in the whole-text condition did seem to attain higher listening comprehension scores to a small degree.

Table 11

Impact of task type on recall scores of each text

Condition	Text	M	SD	t	p	d
Whole-text	Johnson (2006)[a]	23.52	15.85	1.16	0.25	0.23
Segmented		19.96	14.89			
Whole-text	Jacobs (2007)[b]	29.44	19.15	0.15	0.88	0.03
Segmented		30.00	18.60			

Note. N= 100 (50 in each condition). [a] Maximum scores using the weighted rubric = 111. [b] Maximum score = 141. Scoring rubrics, complete with the total number of weighted pausal units possible for recall by each student, can be seen in Appendix B.

The confounding variable "topic familiarity" was investigated in order to determine whether this could have played a role in students' recall scores, and is especially interesting in light of the larger effect size found for the Johnson (2007) text than the Jacobs (2008) text. Table 12 shows that for the four topics rated by students for topic familiarity on a scale from 1-5 (1 = no knowledge of the topic; 5 = expert), no topics had a statistically significant relationship with students' overall recall scores, regardless of condition. When looking at the conditions individually, students' recall scores from the segmented condition had a statistically significant relationship with only one topic ("Outsourcing"). As students' knowledge of outsourcing increased prior to listening to the Jacobs (2007) lecture, their written recall scores also increased. Eight of the fifty-six pausal units (see rubric in Appendix B) revolved around Jacobs' (2007) experience with outsourcing, in his case, his entire life to a group of people in India for a magazine article he wrote. Why the relationship between knowledge of outsourcing and recall scores was seen with the segmented students and not the whole-text students is unknown, although

the topic may have been made more salient for students in the former condition due to the task type.

Table 12

Correlation coefficients of ranked prior knowledge subtopics and recall scores

Lecture	Lecture Sub-topics	Conditions		
		All students	Whole-text	Segmented
Jacobs (2007)	-Outsourcing	0.14	-0.20	0.44*
	-Using the encyclopedia	-0.22	-0.19	-0.33
	-Religions of the world	0.15	0.10	0.17
	-Magazine writing	0.00	0.10	-0.09
Johnson (2006)	-Cholera	-0.09	-0.14	-0.03
	-19th century London	-0.15	0.00	-0.28
	-British monarchs	-0.18	-0.14	-0.23
	-Disease outbreaks	-0.01	0.20	-0.23

Note. N = 60 (30 in each condition). Not all students completed this part of the pre-task survey. * = Correlation is significant at the p = 0.05 level.

Discussing the impact of task type on listening comprehension

Descriptively, there are some interesting differences regarding how task type impacted listening comprehension of students in the whole-text and segmented repetition conditions when listening to different lectures (see Table 11). First, students in the segmented condition received slightly higher recall scores for the Jacobs (2007) lecture than students in the whole-text condition. In addition, students in the whole-text condition attained higher listening comprehension scores when listening to the Johnson (2006)

lecture to a small degree; this is supported by the effect size estimate ($d = 0.23$). Both of these issues may be related to the content and organization of the two lectures (see Appendix B for individual lecture recall rubrics which contain a transcript of each lecture). The segments in the Jacobs (2007) lecture were organized around the different "experiments" he conducted when researching topics for his magazine articles. While the segments were connected to the overarching theme of the lecturer's life being a series of experiments, they were also fairly self-contained. In contrast, the segments in the Johnson (2006) lecture were much more connected to the overarching theme of what led to the cholera outbreak in 1850s London and were not nearly as self-contained as those from the Jacobs (2007) lecture. This raises the issues of whether segmenting a text is more appropriate for certain types of lectures than others, whether segmenting a text might actually harm comprehension when applied to certain types of lectures, and which characteristics are important to take into consideration.

While the descriptive differences are interesting to note, the lack of statistical significance, as well as the absence or small value of effect sizes, indicates that varying the presentation of repeated input to intermediate ESL students does not impact listening comprehension. This helps to explain why different types and styles of repeated input are termed "repetition" in both published research (see Cabrera and Martinez, 2001; Cervantes and Gainer, 1992; Jensen and Vinther, 2003; O'Bryan and Hegelheimer, 2009) and listening textbooks (see Dunkel, Pialorsi, and Kozyrev, 1996; Lebauer, 2000; Salehzadeh, 2006). Repetition is a very powerful task variable; it has been found to have a positive effect on listening comprehension (Cabrera and Martinez, 2001; Cervantes and

Gainer, 1992), and can allow learners more time to process information in the input as well as the relationships between syntactic forms (Hatch, 1983; Jensen and Vinther, 2003; O'Bryan and Hegelheimer, 2009). It is possible that the very act of repeating a listening text could help make main topics salient and assist with structure building independent of how the repetition task is structured. Both of these issues will be investigated further when analyzing the data from the second research question, but what is clear is that segmenting a text into semantic episodes, and combining this with a repetition task, does not have a statistically significant impact on student recall. What is worth noting are two issues that may have impacted the results of the first research question: the role of memory (and forgetting), and task familiarity. A third issue, the ability of the task to hold student attention, will also be discussed in light of informal observations collected during the classroom (quantitative) task and post-task interviews from the think-aloud (qualitative) task.

The role of memory on task performance

It was hypothesized that presenting information in smaller, theme-based chunks would allow for "more structured representation [of the input] in memory and especially better recall" (van Dijk, 1981, p. 191). Therefore, during the procedure, students in the segmented group were presented with one segment that was immediately repeated before moving on to the next segment. Prior research suggested that presenting the repeated input in this way this would allow students to fully process one segment for both meaning and form (Jensen and Vinther, 2003; O'Bryan and Hegelheimer, 2009) before moving on to the next segment. In other words, for each 3-minute text, by the time students in the

segmented condition reach the last word, it has been about 6 minutes since they heard the first word of the text. In contrast, students in the whole-text listened to the entire 3-minute text once, and then again; it was only 3 minutes from the first word to the last. What was not taken into consideration when setting up the tasks was the time difference between the beginning of the text and the end for students in each condition and the impact this difference might have on memory and forgetting.

Researchers in cognitive psychology do not completely agree on what would cause someone to forget information from, for example, the beginning of a 3-minute listening text to the end. There is debate about whether memories fade, or decay, due to the passage of time (i.e. decay theory) or whether the learning of something new as a text progresses (i.e. interference) causes the forgetting of older material. In one empirical example supporting decay theory, Squire (1989) looked at peoples' abilities to recognize the name of a television show for a varying number of years after it had been cancelled. He found that recognition dropped rapidly initially, but then slowed down. The implication for the present study is that because students in the segmented group experienced a greater time delay from the beginning of the text to the end, the decay theory of forgetting may have had an impact on student performance on the written recall in the segmented condition.

Anderson (2000) states that many people "often believe that their memories spontaneously decay with time" (p. 231), but also argues that decay theory cannot be the only theory of forgetting. As time passes following, in this case, the beginning of each 3-

minute lecture, "there is more opportunity to learn new material, which will interfere with retention of old material" (p. 241). In the present study, the amount of possible material learned from the beginning of the lecture to the end would have been the same for students in both conditions, although the task itself may have presented different degrees of *contextual interference*, or interference stemming from a mixture of disconnected information presented within one context, that could have ultimately affected retention. When looking into the retention of paired English-French vocabulary combinations, Schneider, Healy and Bourne (2002) found that when presenting the information to 64 non-French-speaking college students in a blocked (i.e. contextually united) format, such as body parts (e.g., "*dos*, back; *bouche*, mouth; *figure*, face" (p. 423)), or a mixed format (e.g., *"dos*, back; *avion*, airplane; *assiette*, plate" (p. 423)), those in the former condition experienced lower amounts of contextual inference and were therefore able to perform better on an immediate recall test than students who were presented a mixed list of paired words.

The implication of Schneider, Healy and Bourne's (2002) findings for the present study seems to be that presenting information in a contextually united format, as in the case of the segmented repetition task where each segment is united around a theme from the lecture, would lead to increased scores on the written recall. However, this is not consistent with the findings presented in Table 10, and when looking at the tasks, it is clear that the whole-text repetition task is not a "mixed" task in the same way as Schneider, Healy and Bourne's (2002); rather, the contexts are simply not made as salient for students in the latter condition. Thus, it is unclear whether the issues of decay and

interference, and especially contextual interference, may have played a role in the amount of information remembered and recalled by students in both conditions, and to what extent each may have played a role. The role that memory and forgetting play in student performance on listening comprehension tasks, however, is worth noting.

The role of task familiarity on test performance

In addition to decay and/or interference possibly impacting segmented students' performances on their written recalls, task familiarity may also have played a role in helping whole-text students' performances. The listening input presented to students in the whole-text condition was a lecture excerpt played in full, and then played a second time. Again, this is similar to what is often recommended in listening textbooks for students at this level (see Dunkel, Pialorsi, and Kozyrev, 1996; Lebauer, 2000; Salehzadeh, 2006), thus making this type of task more familiar than the task presented to students in the segmented condition. This was confirmed in the post-task interviews where all six of the students participating in the segmented condition stated that they had not listened to an entire listening text in this way prior to the study.

While the notion that more familiar language tasks would result in higher test scores seems logical, very few researchers have investigated the issue empirically. However, there is evidence that this notion has some truth to it. Peña and Quinn (1997) administered a familiar (description) and unfamiliar (one-word labeling) task to fifty-seven Puerto Rican and African American children as part of a routine speech-language screening. What they found was that students performed significantly better on the

familiar test task. Therefore, it is worth considering that differing levels of task familiarity may have impacted student performance on the written recall protocols.

Ability of task to hold student attention

One final reason for the lack of statistical significance found in response to the first research question may have to do with the segmented repetition task itself and its ability to hold student attention. Classroom observations were recorded informally during a limited number of research sessions. What these informal observations revealed was that students in the segmented condition seemed more likely to doodle, chat with friends, or text on their cell phones as the segments were repeated. In contrast, students in the whole-text condition seemed to be more "on task" during the second time as they listened for areas of the text that they felt needed repeated. While this behavior of students in the segmented condition was discouraged by both the researcher and classroom instructor, it indicates that students may not have perceived a need to repeat each segment of the text, whether it be from the text being understandable the first time played, or too difficult to understand in the first place. Therefore, they may not have felt a need to pay attention during the second time each segment was played. Another reason might be related to the unfamiliar format of the task; perhaps students simply lacked the listening strategies needed to successfully comprehend information presented in this unfamiliar format.

Although the classroom observations were not structured and any conjectures stemming from them should be taken with a degree of caution, post-task interviews with students in the segmented condition also shed light on this issue. When students were asked how

they typically repeated texts, many echoed the statement by S025 who said “Anything I can't understand I will repeat it and try to figure out.” However, many also echoed S027’s sentiment that repeating segments was “really helpful I think because if I heard the whole lecture there's much more hard. If you broke it into pieces that's much more helpful. You can just focus on this tiny information. It's much more easier than you listening whole.” While the students may have felt that listening to shorter pieces was helpful and easier than listening to the entire text in full, each segment was repeated whether students felt they needed to listen again or not. This possibly needless repetition of each segment made some students feel the task was tedious. When asked how he felt about repeating the segments, S028 responded “I don’t like it, but I thinks it’s really helpful for listening. [Researcher: “Why do you not like it?”] Because it takes too much time.” In contrast, students from the whole-text condition did not remark on the issue of repetition taking up additional time. Based on this anecdotal observation and post-task interview data, it seems that while students may view the segmented repetition condition as helpful, the task is time consuming and may drive student attention away from the input by repeating information in small increments that does not need repeating. In addition, students may lack the listening strategies needed to successfully comprehend information presented in this unfamiliar format.

Although the difference in listening comprehension was not statistically significant, task type did seem to influence the comprehension problems that students in each group experienced. This helps give additional insight into the impact each type of task had on

students' listening comprehension processes and will be discussed along with the presentation of the results of the second research question.

Research question 2: Impact of task type on listening comprehension problems

Data from twelve participants was used to answer research question 2. In addition to the demographic and overall listening proficiency information provided in Table 2, Table 13 contains information regarding the written recalls scores for participants who completed the verbal protocols and post-task interviews.

Table 13

Participant profiles of students participating in verbal protocols

Student ID	Native Language	Recall score (total)	Mean	SD
W003	Japanese	54	46.5	23.24
W004	Malay	85		
W017	Chinese	36		
W018	Akan (Ghana)	55		
W035	Chinese	22		
W050	Chinese	27		
S020	Malay	52	57.83	23.16
S025	Chinese	85		
S026	Japanese	70		
S027	Chinese	76		
S028	Korean	25		
S029	Korean	39		

In order to answer research question 2 which looks at the impact each task had on the comprehension problems students reported during the think-aloud protocol, it was first necessary to normalize the number of problems reported as students in the segmented group talked for a longer period of time than students in the whole-text group, resulting in

vastly different frequency counts for the number of problems reported. This was primarily due to the way the task was set up; students in the segmented group were given a chance to talk after listening to each segment, each time. In contrast, students in the whole-text group were given a chance to talk after listening to each segment the first time and then had the option of pausing the recording anytime after that. In all, twelve different comprehension problems were reported in the verbal protocols. For students in each condition, Table 14 shows the total number of problems reported, the number of minutes students verbalized in each condition, and the rate of problems reported for each minute students talked (problems per minute). This rate is useful in comparing how frequently each problem occurred among students in each condition.

Table 14

Comprehension Problems per Minute of Time Talked

Condition	Number of Problems	Minutes Talked	Problem Report Rate (Problems per Minute)
Whole Text	98	144.21	0.68
Segmented	124	163.34	0.76

When looking at the problem report rate in Table 14, students in the segmented condition reported slightly more problems than students in the whole-text condition. Again, it is not a surprise given the way the think-aloud task was set up for students in the two conditions. Students in the segmented condition were given more planned time to talk and report their comprehension problems, no matter how slight; students in the whole-text condition, while asked to pause during the second time through the text whenever they had anything to add or clarify from listening the first time, likely did not pause for

every new or continuing problem that they encountered. This is an issue that may be related to training in addition to the actual task itself.

Table 15 shows which problems were reported, the phase of comprehension in which that problem occurred, the individual problem report rate, and the percent value that each problem was reported in relation to the total number of problems reported by students in that condition. The percentage of problems reported in relation to the total number of problems reported by students in that condition provides an opportunity to see how severe a problem was for students in each condition relative to the other problems reported by students in that condition.

The data revealed twelve comprehension problems that occurred during the three phases of students' listening comprehension processes. Seven of these problems occurred during the initial stage of comprehension, or the perceptual stage. According to Anderson (1985), this is the stage where the learner recognizes sounds and segments those sounds into words. Two problems occurred during the parsing stage where listeners assign recognized words into grammatical categories and assign structural and semantic relations. These words are then transformed into a mental representation of the combined meaning of words. Finally, two problems occurred during the utilization stage where learners begin making connections between this newly-parsed information and the knowledge they have about the world.

When looking at the percentage of total problems reported for students in each condition, the majority of the twelve problems were not frequently reported. Therefore, the following discussion will focus on the problems reported in the highest proportion (at least 10%) for students in each condition in order to expand on the results from the previous research question. The problems are presented in the order in which they occurred during the comprehension process, i.e. perceptual problems are discussed first, followed by the perceptual and utilization problems. Finally, additional problems that were not frequently reported, but do show some interesting differences between the groups, will also be discussed.

Cannot chunk streams of speech

Difficulty chunking streams of speech was evident during the perceptual stage of comprehension. This problem was experienced by students in both the segmented and whole-text conditions, although it was reported slightly more often by students who listened to the segmented texts. While the segmented task was structured in order to enhance both attention to the task and structure building (Gernsbacher, 1990), it was not designed to address many of the perceptual processing problems, such as trouble chunking streams of speech, experienced by participants in the study. During the think-aloud procedure, students reported that they were either unable to identify word boundaries in long streams of speech, or they demonstrated an inability to identify "recognisable words or phrases" (Goh, 2000, p. 64). Chunking input is one of the essential decoding processes (Field, 2008) that learners must master before input can go on to be parsed and utilized. One example of a chunking error is seen in S028's verbal

Table 15

Overview of Comprehension Problems

		Whole-Text Condition			Segmented Condition		
Phase of Comprehension	Comprehension Problems	Problems Reported	Problem Report Rate (Problems/ Minute)	Percentage of Total Problems	Problems Reported	Problem Report Rate (Problems/ Minute)	Percentage of Total Problems
Perceptual	Cannot chunk streams of speech	13	0.09	13%	17	0.10	14%
Perceptual	Unknown vocabulary	19	0.13	19%	16	0.10	13%
Perceptual	Do not recognize words they know	1	0.01	1%	1	0.01	1%
Perceptual	Miss beginning of texts	3	0.02	3%	0	0.00	0%
Perceptual	Miss information because of earlier problems	2	0.01	2%	2	0.01	2%
Perceptual	Miss information (reason not specified)	9	0.06	9%	11	0.07	9%
Perceptual	Neglect the next part when thinking of meaning	5	0.03	5%	10	0.06	8%
Parsing	Quickly forget what was heard	2	0.01	2%	6	0.04	5%
Perceptual	Unable to concentrate	5	0.03	5%	4	0.02	3%
Parsing	Unable to form a mental representation	18	0.12	18%	18	0.11	15%
Utilization	Confused about key ideas	0	0.00	0%	2	0.01	2%
Utilization	Understand words but not intended meaning	21	0.15	21%	37	0.23	30%
	TOTAL	98			124		

N = 6 students in the whole-text group; 6 students in the segmented group

protocol when listening to the Jacobs (2007) lecture. The original lecture excerpt is presented, followed by S028's think-aloud.

> My most recent book was uh my previous book was called The Know It All and it was about the year I spent reading the encyclopedia Britannica from A to Z. (Jacobs (2007) lecture)

> I don't know why the speaker's saying about his previous book. Its name is nevida." (S028, segmented condition)

Rather than recognizing the accurate title of the book—The Know It All—S028 instead reports hearing just one word that has no meaning in English—"nevida". Rost (2002) reports that in the perceptual stage, learners use a variety of cues to help segment sounds into words, including identification of phonemes, recognition of word boundaries, and syllable stress. In this example, S028 is failing to use these cues, or at the very least, errs when attempting to make use of them. Because "nevida" is not a recognizable word in English, S028 is unable to parse it in the next phase of comprehension. During the post-task interview, S029 identified the problem of recognizing word boundaries as one she still experiences, although she is getting better at dealing with chunking longer streams of speech because she is more familiar with vocabulary terms:

> Like now I know, Am I able to blah blah blah, but it feels like AmIableto feels like one word. But now I know what does it mean.

> But before, what I mean is that, is also feels like one word. So when some terms that I don't know, they say the terms, and I can't understand. (S029, segmented condition)

As indicated by S029, experience in hearing a common phrase such as "Am I able to" and familiarity with the words contained in the expression are keys to overcoming this problem. In fact, many listening researchers have advocated that instructors spend more time working on close listening exercises that can help address this commonly-occurring decoding problem. For example, Field (2008) advocates the use of short dication exercises to help target frequently-occurring groups of words (e.g., "*AmIableto*"), while Cauldwell (2002) adds that encouraging students to mimic the "features of the acoustic blur that they can hear" (p. 11) can help them become "*familiar and comfortable* with the acoustic blur of normal everyday spontaneous speech" (p. 11, emphasis in italics). Regardless of task type, listeners at this level seem to need more practice decoding input in order to address this perceptual processing problem.

Unknown vocabulary

Not knowing vocabulary was a problem common to students in both conditions, although it made up a greater percentage of the problems reported by students in the whole-text condition (19%) than in the segmented condition (13%). It occurred during the perceptual phase meaning that because of this problem, students were unable to place the information into STS (Nagle and Sanders, 1986), thereby preventing the information from being parsed or utilized. These problems are seen in the examples below:

> In this part I feel many vocabulary I can't understand, so I didn't catch much. (W017, whole-text condition)

> This part there was lot of words that I don't know…I found many words I couldn't understand. New words to me. (W004, whole-text condition)

> I think my vocabulary is a problem. New words, I never see before. (S025, segmented condition)

It was hypothesized that providing repetition in a segmented condition would assist students in attending to the task at hand and encourage the use of more top-down than bottom-up processing, thereby helping them overcome the problem of focusing on words rather than phrases reported by Bacon (1992). In addition, the mere presence of repetition in the tasks was thought to assist students with unknown vocabulary by giving them more time to make inferences about unknown words. However, in these examples students are not focusing on just one or two individual words, and they are not merely "missing" the input the first time. Rather, all of these examples indicate that a number of the vocabulary items were completely unknown. During the post-task interview, S029 (segmented condition) sheds some light on this problem by stating that "If I don't know the terms, I can't understand even second time. But if I missed the terms, second time I will [hear] what I missed the first time." Nation (2001) states that listeners "need at least 95% coverage of the running words in the input in order to gain reasonable comprehension and

to have a reasonable success at guessing from context" (p. 114). While the task structure may facilitate top-down processing, without this crucial vocabulary knowledge, the presentation of repetition may make little to no difference in facilitating comprehension of the text.

Unable to form a mental representation from words heard

Another frequent problem, the inability to form a mental representation of a text, occurred during the parsing stage. Here, students had trouble parsing the syntax or meaning of the input. This problem was directly addressed in the design of the segmented repetition task as both Gernsbacher (1990) and van Dijk (1981) supported the idea that comprehenders would use the initial sentences of paragraphs in a text or, in this case, a segment, to "lay a foundation" (Gernsbacher, 1990, p. 5) for building a mental structure. Once this initial foundation was laid, Gernsbacher claimed that subsequent input would be "mapped onto a developing structure because the more coherent the incoming information is with the previous information, the more likely it is to activate similar memory cells" (p. 2). This process would facilitate the building of a coherent mental structure of the text.

The information in Table 14 supports Gernsbacher (1990) and van Dijk's (1981) claims by showing that the problem of being unable to form a mental representation from words heard was reported slightly more frequently by students in the whole-text condition (0.12 problems/minute) than by students in the segmented condition (0.11 problems/minute), and made up a greater percentage of the problems reported by students in the whole-text condition (18%) than in the segmented condition (15%). An example of this parsing

problem can be seen in the verbal protocol of W050 as he listened to the Johnson (2006) lecture. The original lecture excerpt is as follows:

> It was an amazingly smelly city, um, not just because of the cesspools but also th- the sheer number of livestock in the city would shock people, not just the horses but people had cows in their attics that they would use for milk that they would kindof hoist up there and keep them in the attic until- until literally their milk went out and they died and then they would kindof drag them off, uh, to the boil- you know, the bone-boilers down the street. Um, so, uh, you would just walk around London at this point and just be overwhelmed with this, wi-with this stench. (Johnson (2006) lecture)

W050 reported his thoughts as follows:

> I just can recognize some words, like cow, horse, but I lost the whole information (W050, whole-text condition)

The theme of this segment—presented in the first sentence, as suggested by Gernsbacher (1990) and van Dijk (1981) —is that London is a very smelly city, and then the author goes on to discuss the reasons for the smell. In this verbal protocol, W050 shows that he understood some of the words presented; however, he does not understand the relationship between the ideas presented and how they relate to the bigger topic of London being very smelly. Perhaps because he is in the whole-text condition and the

episodic theme is not made salient, W050 fails to use the theme of the episode (i.e. the smell and reasons for it) to form a mental representation of the input. By focusing on individual vocabulary items rather than finding the connection between the ideas in the text, W050 is unable to form a clear mental representation of the text.

Another example is seen in S028's think aloud following the presentation of the original lecture excerpt:

> So uh, I I work for Esquire magazine. A couple of years ago I wrote an article called My Outsourced Life. Where I hired a team of people in Bangalore, India to live my life for me. So uh they answered my emails, they answered my phone, they argued with my wife for me, uh and they uh uh they read my son bedtime stories. It was the best month of my life. 'Cause I just sat back and I read books and watched movies, uh it was a wonderful experience. (Jacobs (2007) lecture)

> I heard many vocabularies but it is hard to make the main topic with these vocabularies because they don't have the common things between vocabularies. Like his son's bedtime and phone, his wife email (S028, segmented condition)

Here, S028 has picked up one some of the information in the episode, but is unable to find the unifying theme of the episode and, therefore, determine what happened to the

characters in this episode. This is despite the fact that the episodic theme was made salient for this student by presenting it at the beginning of a new segment. Anderson (1985) says that having a mental representation of input is crucial for making inferences and elaborations in the utilization stage. As it is, both W050 and S028 are stuck in the parsing stage with an incomplete understanding of the input.

Despite the fact that students in the segmented condition were provided with salient episodic themes at the beginning of each segment, the parsing problem of being unable to form a mental representation from words heard was experienced by students in both conditions in relatively high numbers. While the episode theme, or macroproposition (van Dijk, 1981), was supposed to help guide students' interpretations of individual word meanings, the mere presence of the macroposition was not enough to encourage these students to combine both bottom-up and top-down processes. One likely reason suggested by Goh (2000) is that the words the students noticed were not key words; rather, they were words that the students were familiar with and could recognize easily. While many nouns were named in the verbal protocols of both students, both failed to recall any verbs that may have given them more of an idea of what was happening in the text and how the ideas were related. One reason these students may not have recognized some of the more "key" vocabulary items in these episodes is that unknown vocabulary was a problem reported in high numbers by students in both conditions. While the task structure may facilitate top-down processing, without this crucial vocabulary knowledge, the presentation of repetition may make little to no difference in facilitating comprehension of the text.

Understand words but not the intended message

Another frequent problem occurred during the utilization phase. Problems encountered during this phase typically revolve around the use of background knowledge to "make [the information] more personally meaningful" (Goh, 2000, p. 57), such as making inferences and elaborating on input. Here, students were successful at understanding the words and forming a mental representation of the input, but failed to make a correct inference or elaboration due a gap in their LTS (Nagle and Sanders, 1986). This problem made up a higher percentage of the problems reported by students in the segmented condition (30%) than by students in the whole-text condition (21%). However, the problem was seen by students in both conditions as shown in the following examples. In the example below, W018 thinks aloud while listening to the Jacobs (2007) lecture. An excerpt of this portion of the lecture follows:

> I'm Jewish in the same way the Olive Garden is Italian. Not very. (Jacobs (2007) lecture)

In the verbal protocol that followed, W018 stated:

> I'm not sure what that phrase or term means to be like the Olive Garden is Italian. I mean, I'm not sure what he's trying to say (W018, whole-text condition)

Here W018 has recognized and understood the words, but he doesn't understand the implication behind the statement. What the speaker is trying to say is that the Olive Garden restaurant calls itself an Italian restaurant, but it might be considered by some (such as the speaker) to serve less-than-authentic Italian cuisine. He makes this point in order to illustrate how Jewish—or not Jewish—he sees himself. W018 does not seem to have the background knowledge necessary to understand the connection. This gap prevents him from relating the new input from the lecture to input stored in his LTS and, therefore, fully understanding the speaker's intent. W018 confirms this finding when asked during the post-task interview whether he felt he experienced any comprehension problems while listening to the lecture: "I understand the literal meaning of the words being said, but I think he's trying to make some points. That I'm not getting it, I'm not getting those points."

While a lack of background knowledge made it hard for some students to understand the implied meaning of a statement, it also made it hard for them to make connections between parts of the lecture, as in the example below stemming from input in the Johnson (2006) lecture.

> Um, they were basically a city living with a modern kindof industrial metropolis with an Elizabethan public infrastructure. Um, so people, for instance, just to gross you out for a second, uh, had had cesspools of human waste in their basement, like a foot to two feet deep. Um they would just kindof throw the buckets down there, uh, and hope that it

> would somehow go away and of course it never really would go away. Uh, and all of this stuff basically had had accumulated to the point where the city was incredibly offensive to just walk around in. (Johnson (2006) lecture)

S029 responded:

> I didn't perfectly understanding about this lecture because he said some Elizabeth public infrastructure. And just people's life. So they just dump, how can I say, dump their waste into their basement. But I didn't understand why it's related with Elizabeth public. (S029, segmented condition)

The background information that S029 is lacking is knowledge about who the Elizabethans are—namely, the British population living during the reign of Queen Elizabeth I (1558-1603)—and knowledge about what the public infrastructure was like during that time. Had S029 been able to access information from her LTS about this period in history, she could have made an inference about what the sanitation conditions were like at this time and, therefore, made the connection between the pieces of information in this segment. Without this background knowledge, however, she was unable to do this.

In discussing the utilization phase of comprehension, Anderson (1985) notes the distinction between suppositions and assertions. In this phase, listeners “try to relate the

information in sentences to knowledge they have about the world…the speaker, in trying to assert new information, must relate it to old information that the listener knows. The speaker is said to *suppose* the old information in order to *assert* the new information" (p. 352, emphasis in original). In W018's verbal protocol example above, the comprehension breakdown occurs because the information the speaker *supposes* is known—that the Olive Garden serves a less-than-authentic version of Italian food—is, in fact, not by this particular listener.

The problem of being able to understand words but not the intended message was experienced by a higher percentage of students in the segment group than by students in the whole-text group. At the outset of the study, it was hoped that the segmented task design would guide learners in identifying the theme of each segment that they could then use to access any prior knowledge of that theme. Also, semantic chunking was also thought to help encourage listeners to focus on semantic cues in the input, thereby activating their semantic memory and encouraging more efficient input processing (Kutas and Federmeier, 2000). However, as in the cases of W018 and S029 above, some students simply lacked the background knowledge needed to fully understand statements in the lectures; this problem cannot be remedied through task design alone.

Other listening problems

The problem of missing information at the beginning of texts was experienced only by students in the whole-text condition (see Table 14). Students in this condition reported

that they “didn’t catch in the beginning” (W017) or “couldn’t hear the first part of this paragraph” (W003). W017 gave some insight into this problem when talking about her comprehension problems during the post-task interview:

> And when I heard the paragraph in the beginning, I can't make my mind to, I don't know how to describe, in the beginning I miss many things because my brain is not can change immediately from the relaxed to heard something. So in the beginning maybe one sentence or one or two sentence I'll miss it. (W017, whole-text condition)

This problem emphasizes the need for repeating input for students at this level of English, as it was also one noted by Goh (2000). However, the fact that this problem was only experienced by students in the whole-text condition suggests that breaking the text into smaller chunks may have made it easier for students to maintain attention to the task at hand as suggested by Robinson (2003) and Samuels and LaBerge (1983).

Another problem occurring during the perceptual phase and making up a slightly higher percentage of problems reported by students in the whole-text condition (5%) than the segmented condition (4%) was the inability to concentrate on the task at hand. Examples from verbal reports include “I just couldn’t concentrate” (W003) and “It’s hard to concentrate” (W050). This is another attention problem that might have been facilitated by the design of the segmented repetition task. However, this conclusion is confounded

by the anecdotal classroom observation and post-task interview data discussed earlier, and highlights the complexity of observing and assessing the issue of attention to task.

Students in the whole-text condition did not experience the majority of every problem, however. In fact, the utilization problem of being confused about key ideas was experienced only by students in the segmented condition. Again, it was hoped that main topics or subtopics of the lecture (i.e., van Dijk's (1981) macropropositions) would be made salient through the segmented task design. However, while this problem was only reported twice—both times by the same student—identifying the main idea or ideas in the lecture was evidently not made clear enough through task design alone. In the post-task interviews, most students reported that this type of segmented repetition was new to them. While the task and the reason for the task design was explained to students in both conditions before completing the quantitative portion of the study (see Chapter 3), perhaps providing students with more pedagogical learner training (Hubbard, 2004) into how the text is segmented and how students can take advantage of the task design would be helpful in overcoming this problem.

Summary

The hypotheses set forth in chapter 2 were that presenting repetition in a segmented format could both enhance recall and facilitate comprehension processes by making input salient, and enhancing both attention and structure building. In response to research

question 1, investigating the impact of task type on listening comprehension, students' written recall scores showed no statistically significant difference between whole-text and segmented repetition tasks. Effect sizes were zero when looking at overall means and the mean scores for the individual lectures, although a small effect size was found for one lecture (Johnson, 2006). The reason for this small effect size was unclear as the students' familiarity self-assessments of four sub-topics from this lecture were not found to be statistically significant confounders. However, given the differences in content and organization between the two lectures, it is possible that segmenting a text is more appropriate for certain types of lectures than for others. While whole-task and segmented repetition task types do not impact listening comprehension scores, informal classroom observation and post-task interview data suggest that while students may view the segmented repetition condition as helpful, the task is time consuming and may diminish student attention by repeating information in small increments that does not need repeating. Finally, the possible roles that memory and task familiarity may play were noted.

Differences between students in each condition were highlighted when looking at the second research question regarding the comprehension problems students in each condition experienced. Of the four problems making up the highest proportion of problems reported, two occurred during the perceptual stage of comprehension. The segmented repetition task was not designed to address many of the perceptual processing problems, such as having trouble chunking streams of speech and unknown vocabulary, experienced by participants in the study. However, both were problems that were

common to learners at this level, and both made up a slightly greater percentage of the problems reported by students in the whole-text condition than in the segmented condition. The ability to form a mental representation from words heard was directly addressed in the design of the segmented repetition task. This problem was reported slightly more frequently by students in the whole-text condition than by students in the segmented condition, and made up a greater percentage of the problems reported by students in the whole-text condition than in the segmented condition. This finding supports Gernsbacher (1990) and van Dijk's (1981) claims that the act of structure building can help students overcome this common problem. The last problem reported in high numbers by students in each condition was concerned with the ability to understand words, but not an intended message. Students in the segmented condition reported this slightly more than students in the whole-text condition, although it was concluded that a lack of background knowledge could not be overcome by task alone.

Finally, evidence from the verbal protocols support the hypothesis that segmenting a text would enhance attention to the task. While neither was reported in high numbers, two problems—missing information at the beginning of texts and an inability to concentrate on the task—were reported either solely by students in the whole-text condition (the former problem) or made up a higher percentage of the problems reported by students in the whole-text vs. segmented condition (the latter). Although this finding does conflict with informal classroom observation and post-task interview data—which must be taken with caution—this suggests that breaking the text into smaller chunks may have made it

easier for students to maintain attention to the task at hand as suggested by Robinson (2003) and Samuels and LaBerge (1983).

The next chapter reiterates the purpose of the study and draws conclusions about whether the way repetition is presented to listeners—either whole-text, or segmented—impacts listening comprehension and comprehension problems. The chapter discusses limitations, and makes suggestions and recommendations in light of the results.

CHAPTER 5: CONCLUSION

The purpose of this study was to explore the impact of two types of listening tasks—the offering of repeated input in either a whole-text or segmented format—on both listening comprehension and comprehension problems in an attempt to determine whether one task type is more effective at helping learners overcome some of the problems common to L2 listeners at the intermediate level and assisting students in reaching a higher level of comprehension. The final chapter summarizes key findings based on quantitative and qualitative evidence and outlines the conclusions that were drawn regarding the impact of task type on L2 listening comprehension and comprehension problems. After the main results are reiterated, the implications for language teachers, materials developers, those interested in the use of technology for language learning and assessment, and future researchers are considered. Finally, the limitations of the study are discussed, and recommendations for future research are made.

Purpose of the study

The purpose of this mixed methods study was to explore the impact of two types of listening tasks, namely the offering of repeated input in either a whole-text or segmented format, on listening comprehension. In addition, the study explored the impact of these task types on the listening comprehension problems students experienced in an attempt to determine whether one task type was more effective at helping learners overcome common problems (e.g., missing information, being unable to form a mental

representation of the text from words heard, etc.) identified in previous literature (see Bacon, 1992; Goh, 2000). The steps taken to answer these research questions, and the findings that ensued, will be summarized below.

Research findings

The research process

The first research question investigated whether task type (whole-text or segmented repetition) would impact listening comprehension when listening to two academic lectures. To answer this question, various quantitative and qualitative data were collected and analyzed.

Listening comprehension was measured using a written recall protocol that was scored using a weighted system based on the relative importance of each pausal unit to the meaning of the passage (see Johnson, 1970). Topic familiarity was considered to be a confounding variable. To assess this, students completed a Likert scale questionnaire asking them to rank their familiarity with twenty topics; eight of these topics were related to the two lectures.

The second research question investigated whether task type (whole-text or segmented repetition) would impact the types of comprehension problems a learner experiences. To answer this question, think aloud data collected from 12 students—6 from each

condition—were used to highlighted differences in how the task affected the ways in which students processed input and the problems they experienced. The verbal protocols were coded for evidence of comprehension problems, i.e., those times when input and knowledge were mismatched. Once coded, each problem was matched to its corresponding phase of comprehension (Anderson, 1985).

Findings

While students in the segmented condition were hypothesized to attain higher scores on the written recalls than students in the whole-text condition, results of a t-test showed that there was no statistically significant difference between overall recall scores (i.e., summed recall scores for the two texts) for students in the two conditions; the effect size was calculated using Cohen's *d* and fell below the small value suggested by Cohen (1988), indicating that a difference in comprehension scores was not present in the sample. When recall scores for each lecture were compared individually, the results mirrored those found when looking at the scores overall in terms of statistical significance, although a small effect size was found for the Johnson (2006) lecture. This indicates that students in the whole-text condition attained higher listening comprehension scores on this text. Topic familiarity did not seem to play a role as for the eight topics on the questionnaire, none correlated with students' recall scores in a way that was statistically significant. When looking at the recall scores of students from each condition, only the topic "Outsourcing" was found to be statistically significant for students in the segmented condition. The reason was unclear, but this topic may have

been made more salient for students in the former condition due to the task type.

Results from the verbal protocol data showed that in total, twelve different comprehension problems were reported; four problems made up at least 10% of the comprehension problems experienced by students in both conditions and were therefore chosen as the focus of the analysis. These included an inability to chunk streams of speech, not knowing vocabulary, an inability to form a mental representation from words heard, and an inability to understand an intended message despite understanding the words heard. The first and last of the four problems made up a higher percentage of the total problems reported by students in the segmented condition than in the whole-text condition, while the opposite was true for the other two problems.

Discussion

It was hypothesized that structuring texts using semantic episodes would help learners create a structured representation of the audio input in memory, which would in turn facilitate recall (van Dijk, 1981). The lack of statistical significance found when analyzing students' written recalls suggests that dividing a text into semantic episodes, and combining this with a repetition task, does not impact recall.

One reason for this finding may be that repetition is a very powerful task variable; it has been found to have a positive effect on listening comprehension (Cabrera and Martinez, 2001; Cervantes and Gainer, 1992) and can allow learners more time to process information in the input as well as the relationships between syntactic forms (Hatch,

1983; Jensen and Vinther, 2003; O'Bryan and Hegelheimer, 2009). It is possible that the very act of repeating a listening text could help make main topics salient and assist with structure building independent of how the repetition task is structured. In addition, decay or interference effects resulting from the structure of the segmented repetition task may have negatively impacted students' abilities to remember information from the beginning of the task to the end. The unfamiliar task of listening to a segmented text that was repeated may have also negatively influenced student performance in this condition.

However what may have also influenced these findings was the impact that the segmented repetition task had on holding student attention. Informal classroom observations during the quantitative portion of the study revealed that students in the segmented condition seemed more likely to doodle, chat with friends, or text on their cell phones as the segments were repeated. In contrast, students in the whole-text condition seemed to be more "on task" during the second time as they listened for areas of the text that they felt needed repeated. These differences, while anecdotal and therefore must be taken with a degree of caution, indicate that students in the former condition may not have perceived a need to repeat each segment of the text, whether it be from the text being understandable the first time played, or too difficult to understand in the first place. One other reason could be that students may have been unable to choose appropriate listening strategies to use when interacting with this unfamiliar text structure. Therefore, they may not have paid attention during the second time each segment was played. Post-task interviews during the qualitative portion of the study with students in the segmented condition revealed that although students felt that listening to shorter pieces of text was

helpful in focusing on the information, they also found the task to be tedious.

While these anecdotal observation and interview data suggest that the segmented repetition task was unsuccessful at holding students' attention during the task, they conflict with those found during the think-aloud procedure. Data from this qualitative portion of the study showed that two problems—missing information at the beginning of texts and an inability to concentrate on the task—were reported either solely by students in the whole-text condition (the former problem) or made up a higher percentage of the problems reported by students in the whole-text vs. segmented condition (the latter). These latter findings suggests that breaking the text into smaller chunks may have made it easier for students to maintain attention to the task at hand as suggested by Robinson (2003) and Samuels and LaBerge (1983).

Findings from the verbal protocol data largely supported the hypotheses that students in the whole-text condition would encounter more problems attending to the task and forming a mental representation of the input compared to students in the segmented condition who were provided with more guidance in attending to the task and structure building through the task's design.

Although problems occurring during the perceptual phase of comprehension, such as trouble chunking streams of speech and unknown vocabulary, and the utilization phase, such as an ability to understand words but not an intended message, were common to learners at this level, the segmented repetition task was not designed to directly address

these perceptual processing problems nor could it assist students with background knowledge that they lacked prior to participating in the study. In contrast, the ability to form a mental representation from words heard *was* directly addressed in the design of the segmented repetition task. This problem was reported slightly more frequently by students in the whole-text condition than by students in the segmented condition, and made up a greater percentage of the problems reported by students in the whole-text condition than in the segmented condition. This finding supports Gernsbacher (1990) and van Dijk's (1981) claims that the act of structure building can help students overcome this common problem.

The results from the study were mixed in terms of supporting and refuting the claims in previous literature. However, both the findings and methods from the study hold a number of implications and recommendations for language teachers, materials developers, those interested in the use of technology for language learning and assessment, and future researchers.

Implications and recommendations

The findings of this research have implications for language teachers, materials developers, those interested in the use of technology for language learning and assessment, and future researchers. First, these findings can help teachers better tailor instruction to their intermediate-level, ESL listening students' needs. L2 listening instructors use a variety of methods to help their students engage with the language

they're learning. However, even when students are obviously struggling with the language and encountering comprehension problems, few have the time or resources to investigate students' real-time listening difficulties. Rather, listening instructors will many times make assumptions about students' comprehension difficulties rather than analyzing learners' needs. For example, a student might incorrectly interpret an utterance (e.g., "nevidea" for "Know it all" as in the case of S028 seen in chapter 4). A teacher might assume that this is a problem with vocabulary since the student replaces a phrase with an incorrect lexical item. Without empirical research, such as the present study, showing that this is actually a problem with chunking streams of speech rather than of unknown vocabulary, this teacher would not know that this kind of perceptual processing problem is one that needs attention. By highlighting some of the issues that L2 listening students struggle with in real time, such as decoding problems during the perceptual processing phrase of comprehension, instructors will naturally be "in a better position to guide our learners in ways of coping with or overcoming some of their listening difficulties" (Goh, 2000, p. 57). These findings can be used to tailor instruction to some of the more common difficulties faced by listeners at the intermediate level.

In addition to shedding light on the real-time listening difficulties that listeners encounter, the study can also help instructors by looking at the ways in which task types encourage or discourage different types of processes and problems. Knowing this, teachers can "help learners develop strategies to compensate for gaps in understanding" (Vandergrift, 2004, p. 10) and order tasks in ways that help learners gradually build up to more automated processing of input. For example, findings from the study showed that the

problem of being unable to form a mental representation of the input made up a greater percentage of problems reported by students in the whole-text condition. Therefore, teachers can build on these findings by training their students to explicitly identify the semantic theme made apparent in each chunk of text in the segmented condition. Once students are comfortable using this grouping strategy (Oxford, 1990) to identify the foundation of the segment (Gernsbacher, 1990), they could begin talking about any prior knowledge they have of the topic and use this knowledge to make inferences about unknown linguistic items or elaborate on the information in the text in order to relate the new information to what they already know. Field (2008) describes a number of classroom activities that help students with this type of structure building. Teachers can then use this segmented task as a type of scaffolding for listening to texts in the whole-text repetition condition by showing students how the same strategies could be used when listening to a text in full. Using "guided activities in the use of…new strategies" (Mendelsohn, 1995, p. 139) will provide students with actual practice in using different strategies while listening; this approach to listening instruction "can help students learn *how* to listen" (Vandergrift, 2004, p. 10, emphasis in original).

While adapting and structuring learning activities is often the job of the language instructor, materials developers are also responsible for understanding learners' needs and then developing appropriate materials. In advocating a strategy-based approach to teaching listening, Mendelsohn (1995) argues that preparatory work is needed before any lessons can be designed. This work includes "devis[ing] ways to help students identify the problems that they are having" (p. 139) through methods such as the verbal protocol.

Being able to draw on findings from the present study, which highlights a number of comprehension problems that could serve as the basis for identifying compensatory strategies that ESL students at the intermediate level need, can also facilitate the job of materials designers and help them design pertinent materials grounded in empirical findings.

Just as there are implications of the present study for teachers and materials developers, there are also implications for the use of technology in language learning and assessment. First, results from the study can help teachers train their students to better handle self-study listening materials delivered via computer or mobile device (e.g., an MP3 player). A plethora of authentic listening materials can be found online and through places like iTunes U, where a variety of academic lectures are recorded (audio or video) and made free to listeners around the world. However, learners typically receive little to no guidance for using these materials in ways that will be most beneficial. The need for teachers to provide training to learners regarding not only the technical aspects of using these types of materials for self study, but also training in the ways in which students learn and process information (see Hubbard's 2004 discussion of learner training), has been documented in recent research (see Hubbard, 2004; Kolaitis, Mahoney, Pomann and Hubbard, 2006; O'Bryan, 2008). With the help of learner training, these students could begin thinking of ways they could structure these electronically-delivered listening passages so as to elicit strategies and facilitate input processing. The present study can help teachers train students who are using self-study materials delivered via computer or mobile device by providing information on the impacts of structuring input on listening

comprehension problems, and by detailing the comprehension processes students encounter while listening.

Secondly, the methods used in the present study have implications for researchers interested in assessing language competence using qualitative data. In the present study, the process (i.e., verbal protocol) data was collected using a digital audio recorder; this data was then imported into a qualitative analysis software program, Transana (Wood and Fassnacht, 2007). While Transana has not been widely explored in the area of language learning and teaching, it offers researchers the opportunity to capture different forms of data and sync them with a written, clickable transcript. In addition to facilitating coding of data, the program also allows researchers to "easily access…analytically significant portions of their [audio and/or video] data" using key words, as well as "view graphical and text-based reports" about their analytic coding (Transana, 2007). This type of qualitative analysis software is invaluable in investigating a wide variety of L2 language issues including listening problems, writing processes, and even pronunciation errors. The use of Transana in the present study provides researchers interested in assessing language processes with an example of how this tool can be used.

Limitations

Two limitations are seen in the present study that can inform future research in this area. First, in order to investigate the impact of repetition task on listening comprehension, students were only asked to listen to two academic lecture excerpts. Although the lectures

were rated as having equal difficulty by a panel of independent raters, the lectures differed in terms of both content and organization. These factors may have contributed to the difference in recall scores on the Johnson (2006) lecture as seen with the small effect size, however with just the two lectures it is difficult to conclude this with certainty. A greater number of lecture excerpts are needed in order to determine whether certain lectures and lecture styles are more appropriate for text segmentation, as well as to determine the lecture characteristics that might influence this.

The second limitation concerns the issue of statistical power. Power is "the probability that [a statistical test] will yield statistically significant results" (Cohen, 1988, p. 1). In this study, power was calculated for the two independent samples t-test used to detect a difference in listening comprehension scores of students in the whole-text and segmented repetition conditions. Power was calculated to be 0.70; this value falls below the 0.80 level suggested by Cohen (1988) that many see as the standard for adequacy and indicates that there is still a 30% chance of accepting a false null hypothesis (i.e., committing a Type II error). Statistical power of a given test depends on three factors: alpha level, effect size, and sample size. While increasing the alpha level to .10 or adjusting the desired effect size could have increased power, both were influenced by recommendations from research methodologists in applied linguistics (Hatch and Lazarton, 1991) and educational measurement (Cohen, 1988). Therefore, the easiest way to increase power in future studies would be to increase the sample size and, therefore, statistical power.

Directions for future research

As seen in the above discussion, the implications of the present study are wide-ranging. It is hoped that research in the area of text segmentation, repetition, and intermediate-level ESL learners' listening comprehension problems will encourage further study by applied linguists. Replication studies would add to the small but growing pool of literature detailing real-time L2 listening comprehension problems experienced by students at the intermediate level and investigating the impact that task has on the problems these students encounter. Doing so can help expand our knowledge of the way language students learn.

Also, replicating this study with students at different levels of English listening proficiency might yield interesting information about the participants for whom the segmented repetition task is most beneficial. Because students in the current study were found to struggle with problems related to perceptual processing, it is possible that providing more advanced listeners with assistance in structure building might impact overall comprehension in a way contrary to what was seen with the participants in the present study.

Other future research could take a strategies-based approach to investigating the differences in student performance between segmented and whole-text repetition tasks. In the present study, no difference was found in the recall scores of students in these two groups. However an investigation of the strategies in which learners engage using both

process and product-based approaches (see Abraham and Vann (1996) and Vandergrift (2007) for suggestions and examples) might yield some interesting findings as to why listening comprehension scores were not impacted by the tasks and how learners worked to build meaning of the input they encountered.

Finally, pursuing the impact that training in listening to segmented, repeated texts might have on listening comprehension could lead to some different, and interesting, results in terms of the strategies they use, problems they encounter, and overall level of comprehension when compared to a whole-text repetition task. Research in this area might help address the comments made by students during the post-task interview who viewed the segmented repetition condition as helpful, but time consuming. Having some training in how to more effectively interact with the segmented repetition task might help them take advantage of the extra time spent listening to the input. In addition, it would help overcome the problem of unfamiliar task negatively affecting student performance observed by Peña and Quinn (1997).

Conclusion

To conclude, this dissertation makes some interesting contributions to the field of Applied Linguistics. It introduces the use of a segmented repetition task, grounded in theories from a range of disciplines, which was found to result in fewer reports of problems common to L2 listeners. It also provides valuable empirical data on the types of problems L2 listeners encounter and ways in which task type influences these problems. Despite a few methodological limitations, the findings and methods from this study have

a number of important implications for teachers, materials developers, and those interested in using technology for language learning and assessment, and are informative for future research.

APPENDIX A. Instructions for Rating Listening Texts

Rating overview

Please award each text a point value from 1-4 for each of the following categories. You are encouraged to use all four points. While you will be rating each text on its own, please also rate it in relation to the other texts.

Overall difficulty

This is an overall rating of difficulty based on your experience with learners at different levels. This is your "gut reaction" to the difficulty of the text. Things you may take into consideration include topic, speech rate, vocabulary, accent, prosodic features, discourse features, and hesitations.

1	2	3	4
Low			High

Perceived content difficulty

Texts at the lower levels will center around a common subject. Listeners will have heard of the topic being discussed, and may have even experienced it. Factors easing a listener's cognitive load including limiting the number of referents, clearly differentiating individuals from objects, including only simple spatial relations and straightforward temporal relations (chronological order is easier than non-chronological), and requiring only a low number of inferences needed to relate one piece of information to the next.

More difficult texts could focus on increasingly abstract concepts that the learners may not have experienced or be able to experience. Factors increasing a listener's cognitive load including an increased number of referents, no clear distinction between individuals and objects, more complex spatial relations and temporal relations (e.g. skipping around), and requiring inferences to relate one piece of information to the next.

1	2	3	4

Low High

Perceived syntactic complexity

Texts at lower levels will present ideas that are chunked into small units. The speaker will avoid using a large number of subordinate and embedded (relative, nominal) clauses.

In more difficult texts, the speaker represents large chunks of ideas through the use of subordinate and embedded clauses.

1	2	3	4

Low High

Discourse structure

Texts at the lower levels will contain obvious rhetorical markers that help signal the major content and sequence in argument. Examples include topic markers (*Let's first deal with…*), asides (*by the way…*), definitions, contrasting, explaining, and concluding. The more difficult texts will lack these rhetorical markers, making it less obvious, and sometimes even difficult, to follow the speaker's main idea(s). In addition, more difficult texts may include "asides" in order to step back from the main content and re-orient the listener (e.g. "he was--this was back in the 1970s--he was a doctor..."), interject humor (e.g. "now what they did--just to gross you out for a second--was they..."), etc. Lower-level listeners may find it difficult to separate these asides from the main content.

1	2	3	4

Low High

Perceived Lexical complexity

For texts at lower levels, the majority of the vocabulary will be frequent words. There should be only a small percentage of academic terms and other, more specialized terms in the text. Of course, if there are technical terms that the lecture spends time defining, explaining, or paraphrasing, then the occurrence of these will contribute less to the lexical complexity of the text than if these words are not defined or explained at all.

More difficult texts will contain more technical vocabulary related to specialized areas. While the majority of the words in the text will continue to be frequently occurring, there should be a larger percentage of academic vocabulary and technical terms specific to the

topic. The lecturer may not spend time defining these terms but instead assume that the listener is familiar with their meanings.

1	2	3	4

Low High

Where does it fall on the scale? (circle)

Were you to place this text along a "difficulty continuum", with the low level being appropriate for IEOP level 4/5 students (the highest level of listening courses offered in IEOP) and the highest being a level appropriate for post-Engl 099L students, where would it fall? Circle a place on the continuum.

|----------|

Low Medium High

Justification/comments (required):

Text 1: “Living Biblically” (Jacobs, 2007)

Episode	Weight	Pausal unit
1	1	I thought I'd tell you a little about
	2	what I like to write.
	3	I like to uh immerse myself in my topics. Uh,
	1	I I just like to uh dive right in and become a
	3	sort of a human guinea pig.
	4	I see my life as a series of experiments.
2	3	So uh, I I work for Esquire magazine, a couple of years ago
	3	I wrote an article called My Outsourced Life.
	3	Where I hired a team of people in Bangalore, India
	4	to live my life for me.
	4	they answered my emails, they answered my phone, they argued with my wife for me,
	3	they read my son bedtime stories.
	3	It was the best month of my life.
	3	Cause I just sat back and I read books and watched movies,
	2	uh it was a wonderful experience.
3	3	More recently I wrote an article for Esquire called
	3	ra- about radical honesty.
	2	And this is a movement where uh
	2	this is started by a psychologist.
	4	In Virginia who says that you should never, ever lie
	1	except maybe during poker and golf.
	1	That's his only, uh, exceptions.
	1	Uh and more than that
	3	whatever's on your brain
	3	should come out of your mouth.
3	2	Uh so I decided I would try this for a month.
	3	This was the worst month of my life. I do not recommend this at all.
	2	To give you a sense of the of the experience uh the article was called
	2	I Think You're Fat.

Episode	Weight	Pausal unit
	2	that was hard.
4	3	my previous book was called The Know It All
	4	and it was about the year I spent reading the encyclopedia
	4	Britannica from A to Z. In my quest to learn everything in the world.
	1	Uh or more precisely from aack,
	1	which is a type of east-asian music
	1	all the way to zyviitch
	1	which is, well I don't want to ruin the ending.
	1	It's a very exciting twist ending like a uh O Henry novel.
	3	But I loved that one because that was an experiment about how much information one human brain could absorb.
	1	although listening to Kevin Kelley you don't have to remember anything you can just uh Google it so
	2	I wasted some time there.
5	3	I I love those experiments but I think that the most profound and life changing
	2	experiment that I've done is my most recent experiment
	4	where I uh spent a year trying to follow all of the
	4	rules of the bible.
	3	The Year of Living Biblically.
6	4	And uh I undertook this for two reasons.
	3	The first was that I grew up
	3	with no religion at all.
	3	I am Jewish in the same way the Olive Garden is Italian.

		2	not very.
			But I've become increasingly interested in religion. I do think it's the defining issue of our time
		4	or one of the,
		1	one of the main ones.
		3	And uh I have a son I want to know what to teach him
6		1	so I decided
		3	to dive in head first and try to live the bible.

Note. The maximum number of pausal units for each weight is as follows: weight 1 = 13; weight 2 = 11; weight 3 = 22; weight 4 = 10.

Text 2: “Cholera” (Johnson, 2006)

Episode	Final		Pausal Unit
	1	1	what I want to do is, is take you back,
		4	to eighteen fifty-four in London,
		3	tell the story, uh, in brief of this outbreak, um, which in many ways
			I think helped create the world that we live in today and and particularly the kind of city
		3	that we live in today.
	2		This period in, in eighteen fifty-four in, in the, you know, middle part of the nineteenth
		2	century in London’s history is incredibly interesting,
		3	London was a city of two and a half million people
			that-on on the face of the planet at that point but it was also the largest city that had ever
		3	been built.
		2	And so the Victorians were trying to kindof
		1	live through and, and simultaneously invent
		4	a whole new scale of living, a scale of living that we, you know, now call
		4	metropolitan living.
	3	1	it was in many ways a- at this point in the mid- eighteen fifties

3 a complete disaster.
they were basically a city living with a modern kindof industrial metropolis with an
3 Elizabethan public infrastructure.
1 so people, for instance,
1 just to gross you out for a second,
4 had had cesspools of human waste in their basement,
3 2 like a foot to two feet deep.
1 they would just kindof
throw the buckets down there, uh, and hope that it would somehow go away and of course
3 it never really would go away.
2 all of this stuff basically had had accumulated to the point where
4 4 the city was incredibly offensive to just walk around in. It was an amazingly smelly city,
not just because of the cesspools but also th- the sheer number of livestock in the city
3 would shock people, not just the horses but
people had cows in their attics that they would use for milk that they would kindof hoist up
2 there
and keep them in the attic until- until literally their milk went out and they died and then
2 they would kindof drag them off,
2 uh, to the boil- you know, the bone-boilers down the street.
5 1 And what ended up happening is that an entire,
3 kindof emerging public health system became convinced that it was the smell that was,
3 that was killing everybody, that was creating these diseases that would kindof
2 wipe through the city every three or four years.
And cholera was really the great killer of this period and it had arrived in London in
4 eighteen thirty-two
3 and every four or five years another epidemic would take
3 ten thousand, twenty thousand people in, in London and, and throughout the UK.
6 And so the authorities became convinced that this, this smell was this problem, we had to
3 get rid of this smell.
2 and so in fact so they, they concocted a couple of early, you know, kindof founding public

		health interventions in the system of the city,
	2	um one of which was called the nuisances act
	3	which they got everybody as far as they could to empty out their cesspools
	4	just poor all that waste into the river.
	3	Because if we get it out of the streets it'll smell much better
	3	and oh right, we drink from the river.
	1	Um, so what ended up happening actually is is
6	4	they ended up increasing the outbreaks of cholera because,
	3	as we now know, cholera is actually in the water, it's a water-borne disease.

Note. The maximum number of pausal units for each weight is as follows: weight 1 = 8; weight 2 = 10; weight 3 = 17; weight 4 = 8

APPENDIX C. Rater Instructions for rating rubrics

In this phase of the study, I am creating rubrics based on 2 lecture transcripts in order to assess student recalls for comprehension. To do this, I'm following a 3-step process:

1. I broke the text into episodes that are based on themes found within the lecture.

2. Next, I broke the text into "pause units" based on places where the speaker pauses. Speakers pause for a variety of reasons—to take a breath, to emphasize something, etc.—and pauses have been shown to have relevance for information processing. The transcripts you receive, then, have already been broken into pause units. Redundancies and infelicities ("um, uh, huh") have been removed.

3. The third step—your job—is to decide which pause units within each episode are important and which are not. To do this, you will be rating each pause unit using a 4-point scale. This will help me assess the quality of students' recalls vs. just quantity.

What I'd like you to do (for each lecture) is to:

1. First either read through the transcript OR listen to the audio version in order to get a general idea of what the lecture is about.

2. Go through the list of pause units for each lecture and begin inserting your weighted rankings in the column next to each pausal unit labeled "Weighting". The weights are explained below.

4 = the most important idea(s) in the episode needed to understand the meaning of the episode;

3 = an important idea in the episode;

2 = a less-important, but still informative idea of the lecture;

1 = non-essential ideas.

Each episode should contain one idea unit that is the most important (ranked "4"), and may even contain two, but not more than two.

APPENDIX D. Pre-task survey

Directions: Please answer all of the following questions to the best of your abilities.

Name: _____

Nationality: _____

Native language(s): _____

Other languages: _____

How long have you been in the U.S.? _____

Age and class (freshman, sophomore, grad, etc.)

_____

Which semester did you take the English Placement Test? _____

University ID# _____

Are you interested in extra listening practice outside of class? ____

Would you be interested in participating in additional research on this topic? _____

On a scale of 1 to 5 (1 = no knowledge at all; 3 = have heard of the topic; 5 = expert on the topic), rate **how much you feel you know** about the following topics by circling the appropriate number.

Topic	Rating scale					Topic	Rating scale				
a. The rainforest	1	2	3	4	5	k. Religions of the world	1	2	3	4	5
b. Cholera	1	2	3	4	5	l. Photojournalism	1	2	3	4	5
c. Apes and language	1	2	3	4	5	m. Using the encyclopedia	1	2	3	4	5
d. Outsourcing services to foreign countries	1	2	3	4	5	n. Disease outbreaks	1	2	3	4	5
e. Designing surveys	1	2	3	4	5	o. Human language development	1	2	3	4	5
f. 19th century London	1	2	3	4	5	p. Perfectionism	1	2	3	4	5
g. Magazine writing	1	2	3	4	5	q. Cultural contexts	1	2	3	4	5
h. Indigenous populations	1	2	3	4	5	r. Child adoption process	1	2	3	4	5
i. British monarchs	1	2	3	4	5	s. Normal distribution (statistics)	1	2	3	4	5
j. Greek philosophers	1	2	3	4	5	t. Refugee camps	1	2	3	4	5

APPENDIX E. Recall protocol prompt

Directions: Please write down everything you felt you understood from the lecture (complete sentences, please). You may write in your native language or in English. Your responses will be assessed based on the number and accuracy of main ideas and details recalled.

APPENDIX F. Recall coding instructions for second coder

Students in the study were asked to listen to 2 lecture excerpts, take notes while listening, and provide a written recall of everything they felt they understood. Your job is to rate each student's recall based on the number of idea units, or pausal units, recalled.

Each listening text has been divided into a number of pausal units; these are units that are divided based on when the speaker pauses. Using the provided rubric of pausal units for each text, you will read the student's recall and give them a score of 0 or 1 **based on whether or not they recalled the main idea of that pausal unit using language from the lecture**. Common problems and examples are given below. Please refer to these often while assessing student recalls.

Problem 1: Recalling parts of idea units

Often, students will recall only part of a unit; if this part is not the main idea of the pausal unit, they should get a score of 0 (see example 1). If this part is the main idea of the pausal unit, they should get a score of 1 (see example 2).

Example 1: Student recall lacking main idea merits a score of zero (0)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"All around me is Italian"	"I'm Jewish in the same way the Olive Garden is Italian."	Student is not getting the main point of this idea unit.

Example 2: Student recall of main idea merits a score of one (1)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"The writer is Jewish the article is written in Italian."	"I'm Jewish in the same way the Olive Garden is Italian."	Student does not get the joke here, but does get the main point that the writer is (a little bit) Jewish

Problem 2: Word boundary and spelling errors in recalls

One other common issue encountered in student recalls is the misspellings of words and incorrect placement of word boundaries in their written recalls. In this case, you must use your knowledge of ESL student phonological errors in order to determine whether they *understand* what the word is even if they cannot identify this word once its heard due to their underdeveloped listening vocabulary. Examples can be seen below:

Example 3: Common word boundary errors in student recalls merits a score of zero (0)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"human gain pick	"sort of like a human guinea pig"	Even though the student is trying to make sense of what he heard, it's just not the same idea at all.

Example 4: Common word boundary errors in student recalls merits a score of one (1)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"His life is serious experiment."	"I see my life as a series of experiments."	Student misses the "series of" part, mistaking this for "serious", but does understand that the lecturer's life is some kind of experiment.

Problem 3: Being too general

Another common error is being too general in the recall. In order to assess whether the student understands what's in each idea unit, you will be looking at the actual words written in the recall and comparing those to the words in the idea unit. This may not always seem fair, as sometimes students will use the language but not necessarily have as deep an understanding as someone who is too general. However it is important that you assess the recall **based on what is written by the student vs. what is implied.**

Examples are below:

Example 5: Being too general merits a score of zero (0)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"But the problem is that everybody threw the trash to their basement"	"had cesspools of human waste in their basement" "throw the buckets [of human waste] down there, uh, and hope that it would somehow go away and of course it never really would go away."	Student doesn't specify what kind of trash is being thrown downstairs

Example 6: Using the same language as the lecturer merits a score of one (1)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"People threw the waste"	"throw the buckets [of human waste] down there, uh, and hope that it would somehow go away and of course it never really would go away."	Student understands that people were trying to get rid of the waste by throwing it away

One final note: remember that you are comparing students' recalls with the individual idea units. You will be **assessing each pausal/idea unit individually**. Even if the student fails to connect one idea unit to the next, they may still identify some main ideas from individual idea units that are worth a score of 1.

Additional examples of when recalls of pausal units would get a score of 0 or 1 cam be seen in the tables below. Please reference these examples often.

Examples of recall data that would be given a score of zero (0)

Student Recall	Pausal Unit from Rubric/Lecture	Reasoning
"human gain pick	"sort of like a human guinea pig"	Even though the student is trying to make sense of what he heard, it's just not the same idea at all.
"All around me is Italian"	"I'm Jewish in the same way the Olive Garden is Italian."	Student is not getting the main point of this idea unit.
"psychology"	"this is started by a psychologist"	Student gets part of this idea unit, but it doesn't really have anything to do with the field of psychology

Examples of recall data that would be given a score of one (1)

Student Recall	Pausal Unit(s) from Rubric/Lecture	Reasoning
"A new living style was developed"	"a whole new scale of living that we, you know, now call"	Student gets the main point about this new way of living
"2.5 million people life in London"	"London was a city of 2.5 million people"	Same words
"Now he writes a book"	"More recently I wrote an article for Esquire called"	Student understands that the lecturer wrote something.
"He uses the encyclopedia to read the book"	"and it was about the year I spent reading the encyclopedia"	Student understands that the author interacted with the encyclopedia in some way
"The writer is Jewish the article is written in Italian."	"I'm Jewish in the same way the Olive Garden is Italian."	Student does not get the joke here, but does get the main point that the writer is (a little bit) Jewish

Problem	Definition	Example
Cannot chunk streams of speech	Words or phrases cannot be chunked into recognizable words or phrases or are chunked into incorrect words or phrases	"When I say one sentence, it's not combination of one words. It feels just like one big word" (S029)
Confused about key ideas	Key points in a message are unable to be identified	"Sometimes in my class after class sometimes panic because I don't know what the professor said was the main point of today's topic" (S028)
Do not know vocabulary	The meaning of a word is not known	"He thinks himself as a human pinikin? I don't know what is pinikin so I have no idea." (S027)
Do not recognize words they know	Words heard are unable to be recognized despite the fact that the word would ordinarily know if seen in its written form.	"And sometimes it's only, I know this word mean, but when I heard it, I can't know it immediately. I need to think a little time. Other times I think it's, I can't, if the word is write it to me, I can know. But maybe my pronounce is wrong. So if heard, it's very different. So this word I can't understand" (W017)
Miss beginning of texts	Beginning of a text or beginning of a segment is missed	"I couldn't hear the first part. Of this paragraph, so I don't know. I was confused" (W003)
Miss information because of earlier problems	Subsequent parts of the input are not understood due to earlier comprehension problems	"Yeah because I don't know what aack, zyviich is that, so it was kindof hard for me to understand the rest of the think about that. So I don't really get what his points." (W004)
Miss information (reason not specified)	Information is simply missed. A reason is not given and sometimes cannot be articulated by the student.	"The last part that I missed it" (S020)
Neglect the next part when thinking about meaning.	Next part of the text is missed when they stop to think about unfamiliar words or the interpretation of a segment of text	"That half of the end I'm not very clear. I think it's the vocabulary before that they explained the Queen Victoria. The vocabulary's a little bit hard to catch so I lost the idea of it." (S027)
Quickly forget what was heard	Information is quickly forgotten.	"When I can't understand a sentence I may forget it very quickly" (W035)
Unable to	Attention is diverted from the input	"I couldn't understand at all. I just couldn't

Unable to concentrate	Attention is diverted from the input	“I couldn’t understand at all. I just couldn’t concentrate.” (W003)
Unable to form a mental representation from words heard	A reasonable mental representation of the input has failed to be derived through the connection of the words heard.	“And I heard, I know he read encyclopedia Britanica, found something on Asians to Zurich, and then how human brain can absorb information. I don't know how those pieces of information can articulate together.” (W018)
Understand words, but not the intended meaning	Literal meaning of the words is understood, but it is difficult to get the full message due to a lack of background knowledge or inappropriate application of	“I’m confused. What is Kevin Kelley? The singer or others? And then he mention you can just google it. I don’t know what is Kevin Kelley so I don’t know why I should google it.” (S027)

APPENDIX H. Instructions for coding comprehension problems

Students in the study were asked to think out loud while listening to 2 lecture excerpts, two times each. They were able to take notes while they were listening. The audio was paused at specific points during the lecture in order to allow each student to tell me what they were thinking: what they understood, what they didn't understand, and how they were making meaning of what they heard. At the end of each lecture, students were asked a few interview questions, including "what were some problems that you had while listening?".

All think-alouds were transcribed verbatim. Your job is to identify all of the listening comprehension problems you find in each student's think-aloud protocol. A **comprehension problem** is *a gap or mismatch between the information conveyed by the speaker, and what the listener understands*. In other words, it's a mismatch between the input and the students' knowledge. Sometimes the listener notices the problem, and sometimes they do not. See table below.

Examples of students' comprehension problems (notice, not notice)

Original Input	Student's Think-aloud
"I think that the most profound and life changing experiment that I've done is my most recent experiment where I uh spent a year trying to follow all of the rules of the bible."	"It's, I don't know, follow road of bible. I don't know that meaning."
"And cholera was really the great killer of this period and it had arrived in London in	"Because of the disease cholera and they take 10,000 people from London to IK

Original Input	Student's Think-aloud
eighteen thirty-two and every four or five years another epidemic would take ten thousand, twenty thousand people in, in London and, and throughout the UK."	because of this disease killing a lot of people…They just took out the people and maybe emigrate to UK."

When you're coding these transcripts, use the established taxonomy created from the training dataset (see table below). When you locate a problem, label it based on what the student identifies the proble as (e.g. missing words, too fast, etc.), or by what is causing the problem (e.g. the student is having trouble making sense of a statement = trouble parsing the meaning of the sentence).

Taxonomy of comprehension problems and examples based on training data

Problem	Definition	Example	Rationale
Understand words, but not the intended meaning	Students can understand the literal meaning of the words, but are unable to get the full message due to their lack of background knowledge or inappropriate application of prior knowledge. (incorrect inferences)goh	"I'm confused. What is Kevin Kelley? The singer or others? And then he mention you can just google it. I don't know what is Kevin Kelley so I don't know why I should google it."	This student understands the words, although he doesn't know who Kevin Kelley is (lacks this background knowledge). The problem is relating the act of "googling" to the person "Kevin Kelley" and trying to figure out how they relate.
Unknown vocabulary	Student does not know the meaning of a word	"He thinks himself as a human pinikin? I don't know what is pinikin so I have no idea."	Student doesn't know what the word means, so he can't figure out the meaning of this sentence.
Neglect the next part when thinking about meaning.	Misses next part of text when they stop to think about unfamiliar words or the	"That half of the end I'm not very clear. I think it's the vocabulary before that	While the student is focusing on the vocabulary, he "loses the idea" of the end

Problem	Definition	Example	Rationale
	interpretation of a segment of text	they explained the Queen Victoria. The vocabulary's a little bit hard to catch so I lost the idea of it."	of the speaker's sentence.
Unable to form a mental representation from words heard	Fail to derive a reasonable mental representation of the input by connecting the words they heard. May be an absence of key words.	"I catched some words. But I couldn't catch the meaning."	Student can't find meaning from the words he heard
Cannot chunk streams of speech	Cannot chunk words or phrases into recognizable words or phrases OR chunk speech into incorrect words or phrases	"It's very hard for me to recognize the words and understanding speech is still low. [Me: understanding words where it's a long stream of speech and it's hard to figure out where the words are?] Yes, yes."	Student agrees that the problem is recognizing words in the stream of speech
Unable to concentrate	Attention is diverted from the input	"I couldn't understand at all. I just couldn't concentrate."	Student says he couldn't concentrate on the input.

REFERENCES

Abraham, R.G., & Vann, R.J. (1996b). Using task products to assess second language learning processes. *Applied Language Learning,* 7, 61-89.

Alderson, J. C. (2000). *Assessing reading*. New York: Cambridge University Press.

Anderson, J. R. (1985). *Cognitive psychology and its implications*. 2nd ed. New York, NY: W. H. Freeman and Company.

Anderson, J. (2000). *Learning and memory: An integrated approach.* 2nd ed. Hoboken, HJ: Wiley.

Bacon, S. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *Modern Language Journal, 76*(2), 160-178.

Bernhardt, E. (1983). Testing foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis, 16*(1), 27-33.

Berne, J. (2004). Listening comprehension strategies: A review of the literature. *Foreign Language Annals, 37*(4), 521-531.

Bremer, K. (1983). *Achieving understanding*. London/New York: Longman.

Bremer, K. & Simonot, M. (1996). Preventing problems of understanding. In K. Bremer, C. Roberts, M. Vasseur, M. Simonot, and P. Broeder (Eds.), *Achieving understanding: Discourse in intercultural encounters* (pp. 159-180). London, UK: Longman.

Brown, G. (1995). Dimensions of difficulty in listening comprehension. In D. Mendelsohn and J. Rubin (Eds). *A guide for the teaching of second language listening* (pp. 59-73). San Diego, CA: Dominie Press, Inc.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Cabrera, M. & Martinez, P. (2001). The effects of repetition, comprehension checks, and gestures, on primary school children in an EFL situation. *ELT Journal, 53*, pp. 281- 288.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19*(4), 727-752.

Cauldwell, R. (2002). Phonology for listening: Relishing the messy. Retrieved April 15, 2010, from http://www.speechinaction.net/SPARC_Richards_publications.htm

Cervantes, R. & Gainer, G. (1992). The effects of syntactic simplification and repetition on listening comprehension. *TESOL Quarterly, 26*, pp. 767-770.

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics, 25* pp. 112-130.

Chamot, A. U. & P. El-Dinary. (1999). Children's learning strategies in immersion classrooms. *The Modern Language Journal, 83*, pp. 319-341.

Chang, Y. F. (2006). On the use of the immediate recall task as a measure of second language reading comprehension. *Language Testing, 23*(4), 520-543.

Cobb, T. (2001). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review, 57*(2), 295-324.

Cobb, T. (2006). *VocabProfile Home*. Retrieved May 26, 2009, from http://www.lextutor.ca/vp/

Cohen, A.D. (1996a). *Strategies in Learning and Using a Second Language*. Longman, NY.

Cohen, A.D. (1996b). Second language learning and use strategies: Clarifying the issues (Revised Version). Paper presented at the Symposium on Strategies of Language Learning and Use – December 13-16, 1994. Seville, Spain.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conrad, L. (1983). Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Acquisition, 7*, 59-72.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle Creek, NJ: Pearson Education

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage Publications.

Creswell, J. W. & Plano Clark, V. L. (2007). *Designing and conducted mixed methods research*. Thousand Oaks, CA: Sage Publications.

Crookes, G. (1991). Power, effect size, and second language research: Another researcher comments. *TESOL Quarterly, 25*(4), 762-765.

Deville, C. and Chalhoub-Deville, M. (1993). Modified scoring, traditional item analysis and Sato's caution index used to investigate the reading recall protocol. *Language Testing, 10*(2), 117-131.

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly, 25*(3), 431-457.

Dunkel, P. & Davis, J. N. (1994). The effects of rhetorical signalling cues on recall. In J. Flowerdew (ed.), *Academic listening: research perspectives* (pp. 55-74). Cambridge: Cambridge University Press.

Dunkel, P., Pialorsi, F. & Kozyrev, J. (1996). *Advanced listening comprehension*. Pacific Grove, CA: Heinle and Helinle Publishers.

Ellis, N. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. Long (Eds.) *The handbook of second language acquisition* (pp. 63-103). Oxford: Blackwell Publishing.

Ellis, N. C. & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly journal of experimental psychology, 49*(A), 234-250.

Ericsson, K.A. & Simon, H. (1984). *Protocol analysis*. Cambridge, MA: MIT Press.

Faerch, C. & G. Kasper (1986). The role of comprehension in second-language learning. *Applied Linguistics, 7*(3), pp. 257-274.

Field, J. (2004). *Psycholinguistics: The key concepts*. New York, NY: Routledge.

Field, J. (2008). *Listening in the second language classroom*. Cambridge: Cambridge University Press.

Flowerdew, J. (ed.), *Academic listening: research perspectives*. Cambridge: Cambridge University Press.

Garson, G. D. (2008). Correlation. Retrieved March 11, 2010, from http://faculty.chass.ncsu.edu/garson/PA765/correl.htm

Gass, S. (2001). Innovations in second language research methods. *Annual Review of Applied Linguistics, 21*, 221-232.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Earlbaum.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, pp. 55-75.

Goh, C. (2002). Exploring listening comprehension tactics and their interaction patterns. *System, 30*, 185-206.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System, 34,* 165-182

Graham, S., Santos, D. & Vanderplank, R. (2008). Listening comprehension strategy use: A longitudinal exploration. *System, 36*, 52-68.

Grgurovic, M. & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language learning and technology, 11*(1), 45-66.

Hahn, L. (1999). *Native speakers' reaction to non-native stress in English discourse*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Hatch, E. (1983). *Psycholinguistics: a second language perspective*. Rowley, MA: Newbury House Publishers.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Los Angeles: Newbury House Publishers.

Hegelheimer, V., & Tower, D. (2004). Using CALL in the classroom: Analyzing student interactions in an authentic classroom. *System*, 32, 185-205.

Hubbard, P. (2004). Learner training for effective use of CALL. In S. Fotos & C. Browne (Eds.), *New Perspectives in CALL for Second Language Classrooms* (pp. 45-67). Mahwah, NJ: Erlbaum.

Jacobs, A. J. (Speaker). (2007). *A. J. Jacobs' year of living biblically*. Retrieved August 12, 2009, from http://www.ted.com/talks/lang/eng/a_j_jacobs_year_of_living_biblically.html

Jensen, E. & Vinther, T. 2003. Exact repetition as input enhancement in second language acquisition. *Language Learning, 53*, 373-428.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33* (7) 14-26.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior, 9*(1), 12-20.

Johnson, S. (Speaker). (2006). *Steven Johnson tours the ghost map*. Retrieved August 12, 2009, from http://www.ted.com/talks/lang/eng/steven_johnson_tours_the_ghost_map.html

Khaldieh, S. (2000). Learning strategies and writing processes of proficient vs. less-proficient learners of Arabic. *Foreign Language Annals, 33*, pp. 522-533.

Kintsch. W. & van Dijk, T. (1978). Toward a model of discourse comprehension and production. *Psychological review, 85*, 363-394.

Krashen, S. (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay and M. Finocchiaro (Eds.), *Viewpoints on English as a Second Language* (pp. 152-161). New York: Regents.

Krashen, S. (1982). *Principals and practice in second language acquisition*. Oxford: Pergamon.

Kutas, M. & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Laufer, B. & Nation, P. (1995). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Lebauer, R. S. (2000). *Learn to listen, listen to learn: Academic listening and note-taking*. 2nd ed. White plains, NY: Pearson Education.

Mackey, A. & S. Gass. 2005. *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.

McLaughlin, B., Rossman, T. & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33*, 135-158.

Meyer, B. (1975). *The organization of prose and its effect on memory*. Amsterdam: North Holland Publishing Company.

Nagle, S. & Sanders, S. (1986). Comprehension theory and second language pedagogy. *TESOL Quarterly, 20*(1), 9-26.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

O'Bryan, A. & Hegelheimer, V. (2009). Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics, 12*(1), 9-38.

O'Malley, J. M. & A. U. Chamot. (1990). *Learning strategies in second language acquisition.* Cambridge, England: Cambridge University Press.

O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics, 10*, 418-437.

Ortega and Iberri-Shea (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual review of applied linguitics, 25*, 26-45.

Peña, E. D. and Quinn, R. (1997). Task Familiarity: Effects on the Test Performance of Puerto Rican and African American Children. *Language, Speech, and Hearing Services in Schools, 28*, 323-332.

Peterson, P. (2001). Skills and strategies for proficiency listening. In M. Celce-Murcia (ed.), *Teaching English as a second or foreign language*. 3rd ed. Boston: Heinle and Heinle, pp. 87-100.

Pujolà, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, 14 (2), 235-262.

Riley, G. & Lee, J. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing, 13*, 173-189.

Robinson, P. (2003). Attention and Memory during SLA. In C. Doughty & M. Long (Eds.) *The handbook of second language acquisition* (pp. 631-678). Oxford: Blackwell Publishing.

Rost, M. (1994). On-line summaries as representations of lecture understanding. In J. Flowerdew (ed.), *Academic listening: research perspectives* (pp. 93-127). Cambridge: Cambridge University Press.

Rost, M. (2002). *Teaching and researching listening*. Harlow: Pearson Education.

Rost, M. (2005). L2 Listening. In Hinkel, E. (ed.), *Handbook of research in second language teaching and learning.* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 503-527.

Salehzadeh, J. (2006). *Academic listening strategies: A guide to understanding lectures.* Ann Arbor, MI: The University of Michigan Press.

Samuels, S. J. & LaBerge, D. (1983). A theory of automaticitiy in reading: Looking back: A retrospective analysis of the LaBerge-Samuels reading models. In L. M. Gentile, M. L. Kamil, and J. S. Blanchard (Eds.), *Reading research revisited*. Columbus, OH: Merrill.

Schmidt-Rinehart, B. (1994). The effects of topic familiarity on second language listening comprehension. *The Modern Language Journal,* 78(2), 179-189.

Schneider, V., Healy, A., and Bourne, L. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46*(2), 419-440.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Shrout, P. & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *2*, 420-428.

Smagorinsky, P. (2001). Rethinking protocol analysis from a cultural perspective. *Annual review of applied linguistics, 21*, 233-245.

Squire , L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology Memory and Cognition, 15*(2), 241-245.

Ur, P. (1984).Teaching Listening Comprehension. Cambridge, UK: Cambridge University Press.

van Dijk, T. (1981). Episodes as units of discourse analysis. In Tannen, D. (Ed.), *Analyzing discourse: Text and talk*. (pp. 177-195). Georgetown: Georgetown University Press, 1981.

van Dijk, T. & Kintsch, W. (1977). Cognitive psychology and discourse. In W. U. Dressler (Ed.) *Current trends in text linguistics* (pp. 61-80). Berlin/New York: de Gruyter.

van Dijk, T. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vandergrift, L. (1997). The strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals, 30*, 387-409.

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled language listener. *Language Learning, 53*, pp. 463-496.

Vandergrift, L. (2007). Listening comprehension in L2/FL learning. *Language Teaching, 40* pp. 191-210.

Vandergrift, L. and Tafaghodari, M. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 1-28.

VanPatten, B. (2004). Input processing in second language acquisition. In VanPatten, B. (Ed.) (pp. 5-31). *Processing Instruction: Theory, research and commentary*. Mahwah: NJ: Lawrence Erlbaum Associates.

Wenden, A. L. (1991). *Learner strategies for learner autonomy.* Englewood Cliffs, NJ: Prentice-Hall.

Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics, 19*, pp. 515-537.

Woods, D. & Fassnacht, C. (2007). Transana v2.2x [Computer software]. Madison, WI: The Board of Regents of the University of Wisconsin System. Available at http://www.transana.org

Young, M. Y. C. (1997). A serial ordering of listening comprehension strategies used by advanced EFL learners in Hong Kong. *Asian Journal of English language Teaching, 7*, 35-53.